

UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

FACULTADE DE MEDICINA

INSTITUTO DE MEDICINA LEGAL



**ESTUDIO DEL GENOMA MITOCONDRIAL EN LA ENFERMEDAD  
MULTIFACTORIAL Y OTRAS APLICACIONES POBLACIONALES Y  
FORENSES**

Memoria que presenta, para optar al grado de doctor,

**Ana Mosquera Miguel**

Santiago de Compostela, marzo de 2010





El Doctor D. Antonio Salas Ellacuriaga, Profesor de la Universidad de Santiago de Compostela y el Doctor D. Ángel Carracedo Álvarez, Catedrático de Medicina Legal de la Facultad de Medicina de la Universidad de Santiago de Compostela,

CERTIFICAN:

Que la presente memoria que lleva por título **“Estudio del genoma mitocondrial en la enfermedad multifactorial y otras aplicaciones poblacionales y forenses”**, de la licenciada en Biología por la Universidad de Santiago de Compostela *Ana Mosquera Miguel*, ha sido realizada bajo nuestra dirección, considerándola en condiciones para optar al Grado de Doctor y autorizándola para su presentación ante el Tribunal correspondiente.

Y para que así conste, firmamos la presente en Santiago de Compostela, a 17 de marzo de 2010.

Prof. Dr. Antonio Salas Ellacuriaga

Prof. Dr. Ángel Carracedo Álvarez

Ana Mosquera Miguel



## *Agradecimientos*

*Es cierto que ésta es una de las partes más difíciles de escribir de la tesis, a pesar de que la mayoría de las personas que aquí menciono ya saben cuánto les agradezco el apoyo que me han dado de una u otra forma.*

*En primer lugar, quiero darles las gracias a mis directores de tesis, Ángel Carracedo y Antonio Salas. Gracias Ángel por darme la oportunidad de entrar en tu grupo, sin ella nada de esto sería posible. Gracias Toño por permitirme trabajar y aprender contigo y adentrarme en el no tan pequeño mundo de las mitocondrias, aunque con ello me haya convertido un poco en “sheriff”.*

*También quiero agradecerles a Maviki y a Chris las colaboraciones y nuevas oportunidades que me han ofrecido.*

*De mis compañeros, a todos les debo algo porque durante estos cuatro años todos y cada uno de ellos me han ayudado de alguna forma. Vane se merece un agradecimiento especial, por lo que me ha enseñado, por los maratones que hemos corrido juntas (nunca literalmente, claro), y por los buenos ratos, fuera y dentro del laboratorio. A Raquel y Fonde, que me enseñaron el laboratorio y los procesos iniciales, surgiendo algún contratiempo que siempre es divertido recordar. Quiero agradecer también a Meli, Paula y Eva, porque para mí son un gran apoyo en el laboratorio. En general, gracias a todos: Mari, Alex, Nuria, Francesca, María Cerezo, Manu, Alberto, Laura Fachal, Luis, Ana Aradas, Cata, Yarimar, Olalla, Ana Pastoriza, Miguel, Carla,*

*Bea, Montse, Rocío, María Dosil, Ángela, Esther, Jens, Ana Ignacio, Paula y Danel. ¡Espero no olvidarme de nadie!*

*Aunque aparentemente mis amigos no tengan mucho que ver con la realización de la tesis, sí que me han echado algún que otro cable informático; pero sobre todo me han ayudado a salir del laboratorio y olvidarme de mitocondrias, ADNmt y pipetas, que también es importante. Por eso, muchas gracias. En especial quiero agradecer el apoyo que me han dado mis compañeras de piso, Patri y Rocío. Sobre todo Rocío porque, además de sufrirme más tiempo, ha aguantado el “making of” de la tesis.*

*A mi familia, que siempre se preocupa de cómo va el CSI de Santiago, y que confían en mí incondicionalmente. En especial, gracias a mis primas, Laura y Marta.*

*A mis padres, porque siempre estáis apoyándome en cada paso que doy, sin dejarme retroceder y dándome el empujón necesario para seguir hacia delante. Gracias por poder contar con vosotros para todo.*

*A mi hermano porque, aunque sea basándose en el desconocimiento, sigue creyendo que me acompañará a Estocolmo. ¡Gracias!*

*GRACIAS A TODOS*

A MIS PADRES





# ÍNDICE

<b>ABREVIATURAS .....</b>	<b>XIII</b>
<b>INTRODUCCIÓN .....</b>	<b>1</b>
<b>1 La mitocondria: aspectos generales .....</b>	<b>3</b>
1.1 Estructura de las mitocondrias .....	3
1.2 Origen endosimbionte de las mitocondrias .....	4
1.3 Biogénesis mitocondrial y segregación mitótica.....	6
1.4 Funciones de la mitocondria .....	6
<b>2 El genoma mitocondrial.....</b>	<b>12</b>
2.1 Estructura del ADN mitocondrial.....	12
2.2 Herencia materna y ausencia de recombinación .....	14
2.3 Tasa de mutación del genoma mitocondrial.....	16
2.4 Variabilidad del ADN mitocondrial .....	19
2.5 Número de copias y heteroplasmía.....	20
<b>3 Análisis de la variabilidad del ADN mitocondrial.....</b>	<b>22</b>
3.1 Evolución de la metodología de genotipado del ADNmt.....	22
3.2 Nomenclatura.....	27
<b>4 Aplicaciones poblacionales de la variabilidad del genoma mitocondrial.....</b>	<b>30</b>
4.1 Generalidades .....	30
4.2 Metodología de análisis.....	31
4.3 Principales aportaciones del ADNmt en genética de poblaciones.....	32
4.4 Limitaciones del ADNmt .....	36
<b>5 Aplicaciones forenses de la variabilidad del genoma mitocondrial ....</b>	<b>37</b>
5.1 Generalidades .....	37
5.2 Tipo de marcadores genéticos.....	38
5.3 Utilidad y limitaciones del ADNmt.....	39

<b>6</b>	<b>Aplicaciones clínicas de la variabilidad del genoma mitocondrial .....</b>	<b>41</b>
6.1	Enfermedades mitocondriales .....	41
6.2	Estudios de asociación caso-control .....	45
	<b>JUSTIFICACIÓN Y OBJETIVOS.....</b>	<b>61</b>
	<b>MATERIAL Y MÉTODOS.....</b>	<b>67</b>
<b>1</b>	<b>Muestras .....</b>	<b>69</b>
<b>2</b>	<b>Extracción de ADN.....</b>	<b>70</b>
2.1	Extracción de ADN a partir de mancha de sangre: fenol-cloroformo .....	70
2.2	Extracción de ADN a partir de sangre .....	71
2.3	Extracción de ADN a partir de otros vestigios biológicos.....	71
<b>3</b>	<b>Genotipado de SNPs de la región codificante mediante minisecuencia.....</b>	<b>72</b>
3.1	Diseño de <i>primers</i> para una PCR multiplex.....	72
3.2	Reacción en cadena de la polimerasa (PCR) .....	74
3.3	Purificación enzimática del producto de PCR: ExoSAP-IT®.....	81
3.4	Reacción de minisecuencia .....	81
3.5	Purificación enzimática del producto de minisecuencia: SAP .....	89
3.6	Electroforesis capilar y análisis de resultados .....	89
<b>4</b>	<b>Secuenciación de la región control del ADNmt.....</b>	<b>90</b>
4.1	Diseño de <i>primers</i> .....	90
4.2	Amplificación de la región control.....	91
4.3	Purificación post-PCR .....	93
4.4	Reacción de secuenciación.....	93
4.5	Purificación post-secuenciación .....	94
4.6	Electroforesis capilar y análisis de resultados.....	96
<b>5</b>	<b>Secuenciación de genomas mitocondriales completos.....</b>	<b>96</b>
5.1	Amplificación de fragmentos solapantes .....	96

5.2	Purificación del producto de PCR en gel de agarosa.....	98
5.3	Reacción de secuenciación.....	98
5.4	Purificación post-secuenciación y electroforesis capilar.....	99
<b>6</b>	<b>Análisis estadísticos .....</b>	<b>100</b>
6.1	Contraste de hipótesis: Test $\chi^2$ .....	100
6.2	Cálculo de Odds Ratio (OR).....	101
6.3	Corrección por test múltiple .....	102
6.4	Análisis de supervivencia: log Rank.....	103
6.5	Regresión logística .....	103
6.6	Análisis filogenéticos.....	104
6.7	Otros análisis realizados.....	105
	<b>RESULTADOS.....</b>	<b>109</b>
	<b>Artículo 1: Is Mitochondrial DNA Variation Associated with Sporadic Breast Cancer Risk? .....</b>	<b>111</b>
	<b>Artículo 2: New Population and Phylogenetic Features of the Internal Variation within Mitochondrial DNA Macro-Haplogroup R0 .....</b>	<b>115</b>
	<b>Artículo 3: Testing the performance of mtSNP minisequencing in forensic samples .....</b>	<b>127</b>
	<b>Artículo 4: Investigating the Role of Mitochondrial Haplogroups in Genetic Predisposition to Meningococcal Disease .....</b>	<b>133</b>
	<b>Artículo 5: Major mtDNA European Haplogroups are not Associated with Schizophrenia .....</b>	<b>143</b>
	<b>Artículo 6: Reassessing the role of mitochondrial DNA mutations in autism .....</b>	<b>159</b>
	<b>DISCUSIÓN .....</b>	<b>175</b>
	<b>CONCLUSIONES.....</b>	<b>181</b>
	<b>BIBLIOGRAFÍA .....</b>	<b>185</b>



*ABREVIATURAS*



- **A:** Adenina
- **AB:** Applied Biosystems
- **ADN:** Ácido desoxirribonucleico
- **ADNmt:** ADN mitocondrial
- **ADNn:** ADN nuclear
- **ADP:** Adenosina difosfato
- **AIF:** *Apoptosis Inducing Factor* (factor de inducción a la apoptosis)
- **ASD:** *Autism Spectrum Disorders* (trastornos del espectro autista)
- **ARN:** Ácido ribonucleico
- **ARNr:** ARN ribosómico
- **ARNt:** ARN transferente
- **ATP:** Adenosina trifosfato
- **BSA:** *Bovine Serum Albumine* (albúmina de suero bovino)
- **C:** Citosina
- **CARD:** *Caspase-Recruitment Domain* (Dominio de reclutamiento de caspasas)
- **CRS:** *Cambridge Reference Sequence* (secuencia de referencia de Cambridge)
- **CSGE:** *Conformational-Sensitive Gel Electrophoresis* (electroforesis en gel sensible a cambios conformacionales)
- **ddNTP:** Di-deoxinucleótido trifosfato
- **dNTP:** Deoxinucleótido trifosfato
- **EDNAP group:** *European DNA Profiling group* (Grupo europeo de perfiles de ADN)
- **ETC:** *Electron transport chain* (cadena de transporte de electrones)
- **FADH<sub>2</sub>:** Forma reducida del dinucleótido flavina adenina (FAD)
- **g:** Gramo
- **G:** Guanina
- **GHEP-ISFG:** Grupo de Habla Española y Portuguesa de la ISFG
- **HVI:** *Hipervariable region I* (región hipervariable I)
- **HVII:** *Hipervariable region II* (región hipervariable II)
- **IAP:** *Inhibitor Apoptosis Protein* (proteína inhibidora de apoptosis)
- **ISFG:** *International Society of Forensic Genetics* (Sociedad Internacional de Genética Forense)

- **kD:** Kilodaltons
- **mL:** Mililitro
- **mM:** Milimolar
- **MRCA:** *Most Recent Common Ancestor* (antepasado común más reciente)
- **NADH:** Forma reducida del dinucleótido nicotinamida adenina (NAD<sup>+</sup>)
- **ng:** Nanogramo
- **nM:** Nanomolar
- **NUMTs:** *Nuclear mitochondrial pseudogenes* (pseudogenes mitocondriales situados en el genoma nuclear)
- **Pb:** Pares de bases
- **PCR:** *Polimerase Chain Reaction* (reacción en cadena de la polimerasa)
- **Pm:** Peso molecular
- **pmol:** Picomol
- **rCRS:** *Revised Cambridge Reference Sequence* (CRS revisada)
- **RFLP:** *Restriction Fragment Length Polymorphism* (Polimorfismo de longitud del fragmento de restricción)
- **r.p.m.:** Revoluciones por minuto
- **SBE:** *Single Base Extension* (extensión de una única base)
- **SDS:** Dodecil sulfato de sodio
- **SNP:** *Single Nucleotide Polymorphism* (polimorfismo de un solo nucleótido)
- **SSCP:** *Single Strand Conformational Polymorphism* (polimorfismo conformacional de cadena única)
- **STR:** *Short Tandem Repeat* (repetición corta en tándem)
- **T:** Timina
- **Tm:** Temperatura de *melting*
- **TMRCA:** *Time to Most Recent Common Ancestor* (tiempo hasta el antepasado común más reciente)
- **µg:** Microgramo
- **µL:** Microlitro
- **µm:** Micrometro
- **µM:** Micromolar



---

## *INTRODUCCIÓN*



## 1 La mitocondria: aspectos generales

La mitocondria es un orgánulo presente en casi todas las células eucariotas que, además de poseer genoma propio, presenta su propia maquinaria de transcripción, traducción y ensamblaje de proteínas. La razón de su ubicuidad hace referencia a la importante función que desempeña, proporcionar el aporte energético necesario para las células. El número de orgánulos que se presentan en cada célula varía en función del tipo celular y de las demandas energéticas de cada momento.

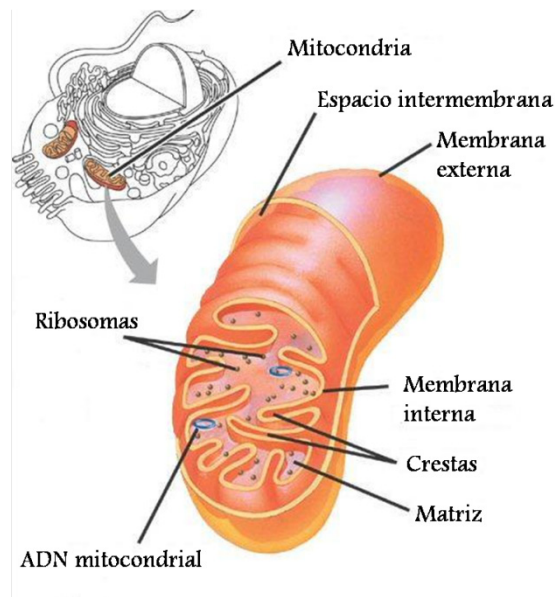
### 1.1 Estructura de las mitocondrias

Las mitocondrias son orgánulos de doble membrana y de forma típicamente alargada, alrededor de 0'5 $\mu$ m de ancho y hasta 10 $\mu$ m de largo, sin embargo pueden variar considerablemente tanto en forma como en tamaño. La **membrana mitocondrial externa** es una bicapa lipídica relativamente sencilla, lisa y muy permeable, gracias a la presencia de numerosas porinas, que permiten el paso inespecífico de moléculas de hasta 10kD de tamaño. De esta forma, iones, nutrientes, ATP, ADP, y otras moléculas, pueden pasar a través de esta membrana con facilidad. Por el contrario, la **membrana mitocondrial interna** solamente es permeable al oxígeno, dióxido de carbono y agua, mientras que el resto de las moléculas necesarias para la mitocondria pasan mediante transporte activo. Además, su estructura es considerablemente más compleja que la de la membrana externa, ya que en ella se localizan proteínas transportadoras, todas las proteínas de la cadena de transporte de electrones y el complejo de la ATP sintasa, esenciales para la función mitocondrial y cuyo número varía dependiendo del tipo celular. Además, la membrana interna también se caracteriza por presentar numerosas invaginaciones que se denominan **crestas mitocondriales**, permitiendo así aumentar su superficie total.

Como consecuencia de la presencia de esta doble bicapa lipídica en el interior de las mitocondrias se diferencian dos espacios. El **espacio intermembrana**, localizado como su nombre indica entre la membrana externa y

la interna, juega un papel importante en la fosforilación oxidativa mediante la creación de un gradiente electroquímico. La **matriz mitocondrial**, de carácter gelatinoso, es el espacio que se encuentra en el interior de la mitocondria, limitado por la membrana interna. En la matriz tienen lugar rutas metabólicas importantes gracias a la presencia de enzimas solubles, cofactores y los substratos necesarios.

Por último, la mitocondria posee varias copias de un **genoma** propio y la maquinaria necesaria para la síntesis proteica (ARNt y ribosomas).



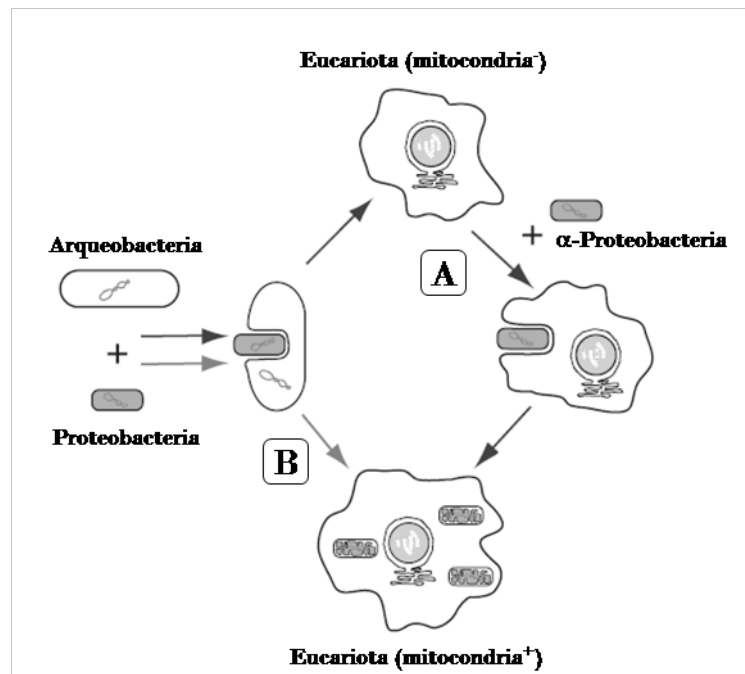
**Figura 1:** Localización y estructura de las mitocondrias.

## 1.2 Origen endosimbionte de las mitocondrias

Las mitocondrias, al igual que los plástidos, presentan muchas semejanzas con las bacterias: su forma, tamaño o la presencia de un genoma circular de doble cadena. Éstas y otras características morfológicas (presencia de doble membrana, químicamente diferente a la nuclear), junto determinadas propiedades bioquímicas y moleculares (no participan en el tráfico vesicular como el resto de orgánulos de membrana simple) y datos paleontológicos llevaron a Lynn Margulis (1976) a formular la **teoría endosimbiótica** o de la **endosimbiosis seriada**. Según esta teoría, el origen de las células eucariotas es consecuencia de la sucesiva incorporación simbiogenética de diferentes procariotas, de forma que los organismos endosimbiontes transfirieron parte o todo su material genético al hospedador.

Hoy en día se acepta que los orgánulos de doble membrana (mitocondrias y plástidos) son el resultado de una o varias simbiosis a partir de las cuales evolucionaron las células eucariotas actuales, adquiriendo así maquinarias bioenergéticas y biosintéticas compartimentalizadas.

Análisis de genes mitocondriales y de su distribución y organización en el genoma de diferentes organismos, parecen confirmar que los genomas mitocondriales han derivado de un único ancestro relacionado con la división  $\alpha$  de las Proteobacterias, hace aproximadamente  $1.5 \times 10^9$  años. Existe mayor controversia con respecto al tipo de célula hospedadora que participó en la simbiosis con la  $\alpha$ -proteobacteria, ya que para los defensores del modelo de endosimbiosis en serie, la célula huésped en ese momento ya presentaba núcleo (resultado de una simbiosis anterior entre una arqueobacteria y una proteobacteria). Sin embargo, la **hipótesis del hidrógeno** propone una simbiosis entre una arqueobacteria metanogénica y  $\alpha$ -proteobacteria metanotrófica, surgiendo el núcleo celular y la mitocondria como resultado de un único evento de simbiosis (Dyall et al. 2004; Gray et al. 1999).



**Figura 2:** Representación esquemática de dos de las hipótesis formuladas para explicar el origen de la célula eucariota, el modelo de endosimbiosis en serie (A) y la hipótesis del hidrógeno (B). Imagen modificada de Gray et al. (1999).

Es posible que filtraciones de ADN durante la división o la lisis de la eubacteria favoreciesen la transferencia de secciones de su genoma al núcleo, y se cree que esto ha contribuido de forma sustancial a la evolución de los genomas eucariotas. Esta redundancia genética podría explicar la posterior reducción del genoma del endosimbionte (Dyall et al. 2004).

Actualmente también tienen lugar transferencias de ADNmt al núcleo, originando secuencias no codificantes denominadas NUMTs (*nuclear mitochondrial pseudogenes*). Sin embargo, parece que la mayoría se originaron a partir de duplicaciones de NUMTs ya existentes en el genoma nuclear (Ramos et al. 2009).

### **1.3 Biogénesis mitocondrial y segregación mitótica**

Las mitocondrias no se sintetizan nunca *de novo*, sino que siempre surgen por crecimiento y división de otras ya existentes. Cada orgánulo ha de duplicar su masa y luego dividirse por la mitad un promedio de una vez por cada generación celular, y para ello han de importar lípidos, nucleótidos y pre-proteínas desde el citosol, de forma que el núcleo juega un papel esencial en la biogénesis mitocondrial (Dyall et al. 2004).

En la mayoría de las células, los orgánulos transformadores de energía se dividen durante la interfase y replican su ADN durante todo el ciclo celular, es decir, que presentan un desfase con respecto a estos mismos procesos en las células. Durante la división celular, las mitocondrias se distribuyen de forma aleatoria entre las células hijas y, como se analizará posteriormente, este fenómeno, denominado segregación mitótica, tendrá implicaciones clínicas importantes.

### **1.4 Funciones de la mitocondria**

#### **1.4.1 Producción de energía celular**

Muchos de los procesos que tienen lugar en las células (metabólicos, biosintéticos,...) son imprescindibles para la supervivencia de las mismas, y requieren aportes de energía para llevar a cabo reacciones endotérmicas. La

mitocondria es un orgánulo celular que sirve como suministro energético a la célula eucariota, ya que oxida determinados sustratos para obtener adenosina trifosfato (ATP). El número de mitocondrias existentes en una célula viene determinado por los requerimientos energéticos propios de cada tipo celular y de cada momento.

Los principales sustratos a partir de los cuales se obtiene energía en la mitocondria son glucosa y ácidos grasos. Es importante asegurar su aporte continuado para el metabolismo oxidativo, y por ello las células animales almacenan ácidos grasos en forma de grasas y glucosa transformándola en glucógeno. Los pasos iniciales del catabolismo, tanto del glucógeno como de las grasas, se llevan a cabo en el citosol, y son sus productos los que son transportados al interior de las mitocondrias, donde tiene lugar la oxidación del piruvato, la  $\beta$ -oxidación de los ácidos grasos, la oxidación de los aminoácidos y el ciclo del ácido cítrico.

La producción de energía a partir de glucosa y oxígeno, mayoritariamente en forma de ATP, es resultado de dos procesos catabólicos. Inicialmente la glucosa es convertida a piruvato por vía glicolítica en el citoplasma, posteriormente es importado a la mitocondria, transformado en acetil-CoA y oxidado hasta obtener dióxido de carbono y agua. Todo esto tiene lugar gracias a dos sistemas interconectados y muy coordinados: el ciclo de los ácidos tricarboxílicos y el complejo de la cadena respiratoria.

El **ciclo de Krebs** (ciclo de los ácidos tricarboxílicos o del ácido cítrico) recibe su nombre en honor a su descubridor, quien, en 1937, propuso que los combustibles orgánicos se oxidaban a través de una ruta cíclica. Así, el ciclo del ácido cítrico consiste en una oxidación progresiva del acetil-CoA mediante un ciclo de ocho reacciones, que tiene lugar en la matriz mitocondrial, durante las cuales se desprende  $\text{CO}_2$ . Sin embargo, su contribución más importante al metabolismo consiste en la extracción de electrones de alta energía durante dicha oxidación. Los principales transportadores de electrones son el NADH y el  $\text{FADH}_2$ , y provienen tanto del ciclo de Krebs como de los pasos anteriores (glucólisis,  $\beta$ -oxidación de ácidos grasos, transformación de piruvato en acetil-CoA).

La **cadena respiratoria** o **sistema de fosforilación oxidativa (OXPHOS)** consiste en el conjunto de reacciones que comprenden la cadena de transporte de electrones y la fosforilación oxidativa. La **cadena de transporte electrónico** (ETC, *Electron Transport Chain*) consiste en la transferencia sucesiva de los electrones de alta energía cedidos por el NADH o FADH<sub>2</sub> a cuatro grandes complejos proteicos, situados en la membrana mitocondrial interna. Cada uno de ellos tiene mayor afinidad por los electrones que su predecesor, hasta que finalmente son transferidos al oxígeno, que presenta una afinidad mayor que cualquier otro complejo de la cadena. Son varios los modelos propuestos para explicar el mecanismo mediante el cual la energía liberada en la respiración se utiliza para impulsar la síntesis de ATP, sin embargo, hoy en día se acepta la hipótesis de **acoplamiento quimiosmótico**, propuesta en 1961 por el bioquímico británico Peter Mitchell y que le valió el Premio Nobel en 1978. Este modelo en su forma más básica propone que la energía del transporte electrónico impulsa un sistema de transporte activo que bombea protones fuera de la matriz mitocondrial al espacio intermembrana, creando así un gradiente electroquímico entre ambos. Los protones del exterior tienen una tendencia termodinámica a volver a entrar en la matriz para igualar el pH (gradiente químico) y la carga (gradiente eléctrico) a ambos lados de la membrana (Mathews et al. 2002). Este flujo de protones a favor de gradiente es utilizado por la ATP sintasa para impulsar la conversión de ADP+P<sub>i</sub> en ATP mediante un proceso denominado **fosforilación oxidativa**.

La cadena respiratoria es la principal fuente de ATP en los organismos aerobios, y se denomina así porque es imprescindible la presencia de oxígeno para que tenga lugar, ya que esta molécula es el aceptor último de los electrones. En la mayor parte de las células, al menos un 90% del oxígeno molecular consumido se utiliza para la fosforilación oxidativa, aunque se conocen otras 200 enzimas que utilizan esta molécula como sustrato. En ausencia de oxígeno la síntesis de ATP se detiene, lo que puede provocar la muerte celular. La evolución de la respiración celular capacitó a los organismos para obtener mucha más energía en comparación con el metabolismo anaerobio. En general, el balance total de energía obtenida mediante la glucólisis anaerobia (transformación de glucosa a piruvato) es de aproximadamente dos moles de ATP, mientras que la oxidación completa de la



glucosa a CO<sub>2</sub> y agua en presencia de oxígeno aporta alrededor de 38 moles de ATP.

#### **1.4.2 Biosíntesis de componentes celulares**

Las principales funciones biosintéticas de la mitocondria provienen de alguno de los intermediarios del ciclo de los ácidos tricarboxílicos e incluyen:

- Biosíntesis de ácidos grasos y esteroides: en algunos tejidos el citrato es transportado al citosol, donde es fragmentado para producir acetil-CoA para la biosíntesis de los ácidos grasos.
- Biosíntesis de aminoácidos, purinas y pirimidinas: a partir del aspartato y glutamato resultantes de la transaminación del oxalacetato y  $\alpha$ -cetoglutarato, respectivamente.
- Hemo y otras porfirinas: a partir del succinil-CoA

Dado que éstas y otras reacciones tienden a causar una reducción de los intermediarios del ciclo del ácido cítrico, el funcionamiento de éste se iría deteriorando si no hubiera otros procesos que repusieran las reservas de los mismos, se denominan rutas anapletóricas. En la mayoría de las células el flujo de salida del carbono del ciclo se equilibra mediante estas reacciones anapletóricas, de manera que las concentraciones intramitocondriales de los intermediarios del ciclo se mantienen constantes a lo largo del tiempo (Mathews et al. 2002).

#### **1.4.3 Apoptosis celular**

La apoptosis, o muerte celular programada, es un programa de suicidio celular mediante el cual células individuales son destruidas mientras que se preserva la integridad del tejido circundante. Esta destrucción celular dirigida es crítica en muchos contextos fisiológicos, como la embriogénesis, maduración y respuesta de células inmunes, homeostasis de los tejidos y respuesta al daño celular. La evasión del programa apoptótico promueve la tumorigénesis.

Uno de los últimos descubrimientos en relación a la biología mitocondrial es el hecho de que este orgánulo juega un papel importante en la apoptosis celular. El complejo papel de la mitocondria en la apoptosis celular de los mamíferos se puso

de manifiesto cuando estudios bioquímicos identificaron varias proteínas mitocondriales capaces de activar directamente programas de apoptosis celular. Normalmente estas proteínas residen en el espacio intermembrana y en respuesta a una variedad de estímulos apoptóticos son liberadas al citosol o al núcleo. Estas proteínas promueven la apoptosis ya sea mediante la activación de caspasas y nucleasas o neutralizando inhibidores citosólicos de este proceso.

El escenario se complica cada vez más, ya que el número de proteínas implicadas en la apoptosis aumenta continuamente, pero se puede generalizar diciendo que la interacción entre proteínas pro-apoptóticas (citosólicas y mitocondriales) y proteínas anti-apoptóticas determinará el futuro de la célula.

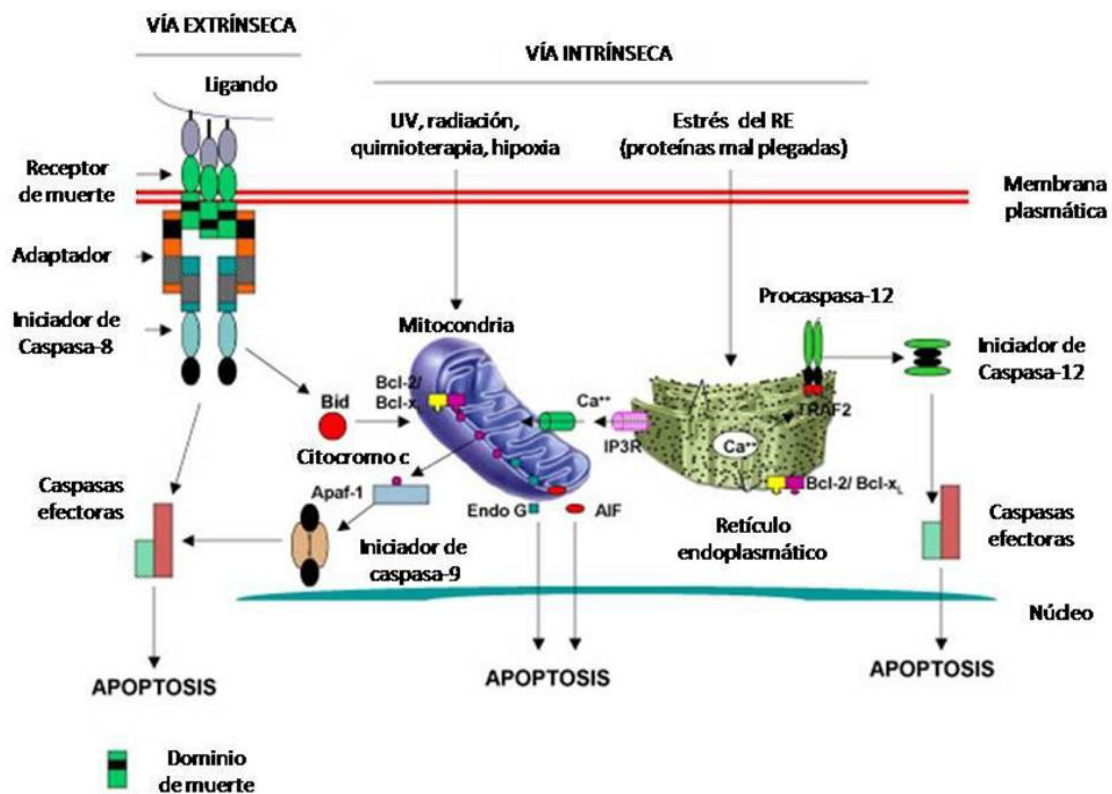
En concreto la mitocondria interviene en la **vía intrínseca** de la apoptosis celular, activada por estímulos como daños en el ADN (quimioterapia o radioterapia), estrés oxidativo o radiación ultravioleta. Estos factores provocarían la permeabilización de las membranas mitocondriales y, como consecuencia, la liberación del citocromo c, uno de los componentes de la cadena de transporte de electrones e inhibidor de proteínas anti-apoptóticas. El citocromo c citosólico se une a Apaf1, una proteína que contiene un dominio de reclutamiento de caspasas (CARD), lo que provoca la oligomerización del complejo, formando el apoptosoma. Esto deja expuesto el dominio CARD, donde une la caspasa 9, y este complejo es capaz de activar la cascada de caspasas ejecutoras, como la caspasa 3 (Wang 2001). También otra proteína mitocondrial codificada por el núcleo, Smac/Diablo, sale al citosol y se une a proteínas IAP (*Inhibitor Apoptosis Protein*), permitiendo así la activación de la caspasa 9 por el apoptosoma.

El proceso se complica aún más ya que existen otras proteínas, AIF (*Apoptosis Inducing Factor*) y endonucleasa G, liberadas también desde el espacio intermembrana de la mitocondria, que se introducen en el núcleo provocando condensación de la cromatina y fragmentación a gran escala del ADN nuclear. En este caso se desencadena un programa apoptótico paralelo a la activación de las caspasas (Wang 2001).

Las caspasas ejecutoras cortan muchos sustratos intracelulares importantes, provocando cambios morfológicos como la condensación de la cromatina,

fragmentación del ADN o rotura de la membrana nuclear. El proceso termina con la formación de cuerpos apoptóticos, fragmentos celulares rodeados de membrana que evitan la salida de todos estos compuestos al exterior y el posible daño a células colindantes. Estos cuerpos apoptóticos son rápidamente fagocitados por macrófagos circulantes evitando la respuesta inflamatoria.

Hay otra vía de activación de la apoptosis, la **vía extrínseca** o mediada por ligando de muerte, en la que intervienen receptores de muerte (TFN, FAS, FADD,...) que se encuentran situados en la membrana citoplasmática y que tras la unión con sus ligandos específicos, llevan directamente a la activación de las caspasas.



**Figura 3:** Esquema de las vías intrínseca y extrínseca de activación de la apoptosis. AIF, *Apoptosis Inducing Factor*; CARD, *Caspase-Recruitment Domain*; IAP, *Inhibitor Apoptosis Protein*.

Con independencia de la ruta de activación, el proceso culmina con la activación de las caspasas efectoras, que junto con otros ejecutores independientes provocan la destrucción celular.

## 2 El genoma mitocondrial

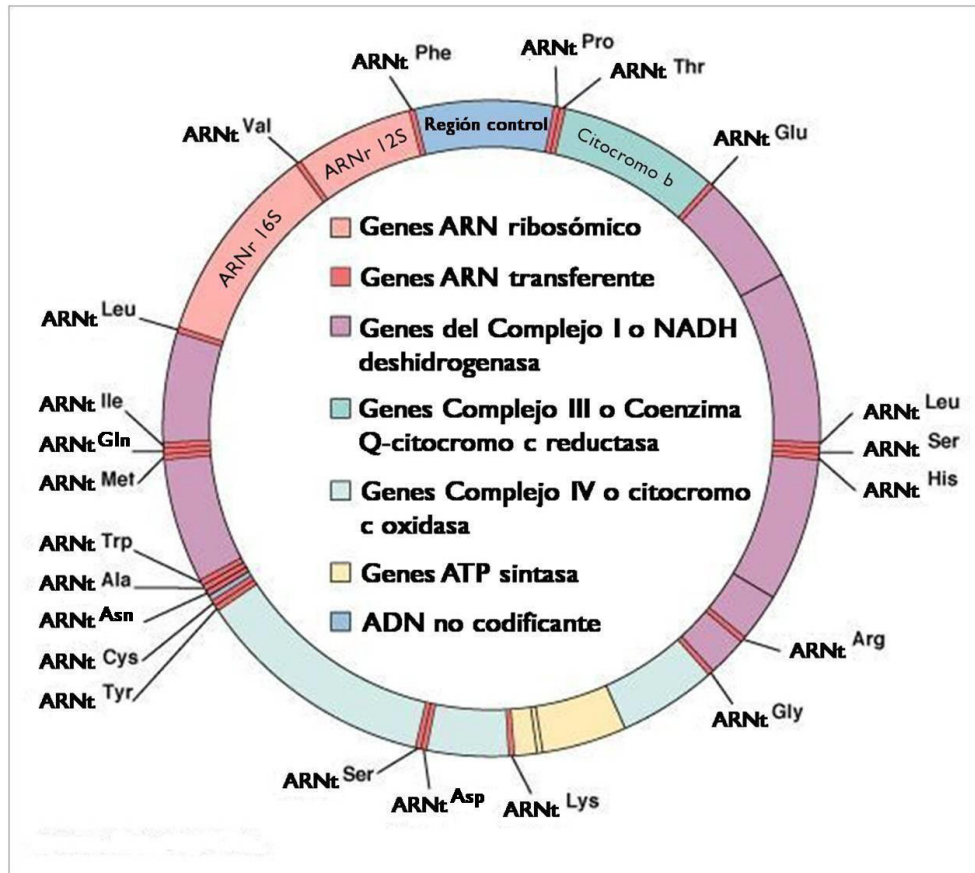
### 2.1 Estructura del ADN mitocondrial

El ADN mitocondrial (ADNmt) humano es una molécula circular de doble hebra, de 16569 pares de bases (pb) de longitud. Las proporciones de los cuatro nucleótidos son desiguales, de hecho, una de las cadenas contiene un mayor número de purinas (G y A), la cadena H (“Heavy”) o pesada, mientras que la otra es rica en pirimidinas (C y T), la cadena L (“Light”) o ligera. Los nombres se debieron a su diferente densidad en cloruro de cesio (CsCl).

Dentro de esta molécula se pueden distinguir a grandes rasgos dos regiones, según la presencia o ausencia de genes. La **región control** no codifica para proteínas, ARNt ni ARNr, pero contiene motivos importantes como los promotores de las hebras L y H, sitios de unión a factores de transcripción o el origen de replicación de la cadena H. El origen de replicación de la cadena L se sitúan entre los genes de ARNt<sup>Cys</sup> (cisteína) y ARNt<sup>Asn</sup> (asparagina). Esta región abarca aproximadamente 1200pb, desde la posición 16024 hasta la base 16569 y desde la 1 hasta la base 576. La **región codificante** consta de 37 genes que codifican para 22 ARNt, dos ARNr y 13 proteínas de la cadena de transporte de electrones, el resto están codificadas por el genoma nuclear. Los complejos proteicos que conforman esta cadena son:

Complejo	Nombre	Constituyentes	Acción
I	NADH-Ubiquinona oxidoreductasa	48 <sup>(a)</sup> polipéptidos (7)	Oxidación NADH.
II	Succinato-Ubiquinona oxidoreductasa	5 polipéptidos (0)	Oxidación sustratos FADH <sub>2</sub> dependientes.
III	Ubiquinol-Citocromo c oxidoreductasa	11 polipéptidos (1;citocromo b)	Oxidación sustratos NADH y FADH <sub>2</sub> dependientes.
IV	Citocromo c oxidasa	13 polipéptidos (3)	Transfiere equivalentes reductores del citocromo c al oxígeno molecular.
V	Adenosina trifosfato sintasa	14 polipéptidos (2;ATPasa 6 y 8)	Convierte el gradiente transmembrana en energía (de ADP a ATP).

**Tabla 1:** Complejos de la cadena respiratoria mitocondrial. Entre paréntesis se refleja el número de péptidos codificados por el genoma mitocondrial. (a) Número máximo de polipéptidos reportado hasta la fecha en la literatura revisada.



**Figura 4:** Representación gráfica de la estructura y organización del ADN mitocondrial humano.

El genoma nuclear codifica para el resto de las proteínas mitocondriales, incluidos todos los polipéptidos del complejo II, las subunidades de la ADN polimerasa  $\gamma$ , los componentes de la ARN polimerasa mitocondrial, el factor de transcripción mitocondrial (mtTFA), las proteínas ribosómicas y las enzimas metabólicas mitocondriales.

En las mitocondrias, el código genético está alterado con respecto al código universal, ya que cuatro de los 64 codones presentan un significado diferente al que posee para la síntesis citoplasmática:

Codón	Código Universal	Código mitocondrial de mamíferos
UGA	STOP	Triptófano
AUA	Isoleucina	Metionina
AGA	Arginina	STOP
AGG		

**Tabla 2:** Alteraciones del código genético del genoma mitocondrial humano.

La observación de que el código genético mitocondrial es casi el mismo en todos los organismos proporciona fuertes evidencias de que todas las células eucariotas evolucionaron a partir de un antepasado común.

## **2.2 Herencia materna y ausencia de recombinación**

Durante muchos años, el hecho de que las mitocondrias, y por tanto el ADNmt, presentaban una herencia estrictamente materna estaba considerado como un dogma incuestionable. Esta característica fue puesta de manifiesto por Giles et al. (1980) cuando, analizando polimorfismos de secuencia mediante RFLPs (*Restriction Fragment Length Polymorphisms*), observó que la progenie siempre mostraba la variante materna.

La explicación se encuentra en el hecho de que, en los animales superiores, el óvulo aporta al cigoto la mayoría o la totalidad del citoplasma (dependiendo del animal), y las pocas mitocondrias que hayan sido aportadas por el espermatozoide son destruidas selectivamente en los lisosomas. Esta destrucción ocurre durante la fertilización y se debe a que estas mitocondrias son marcadas durante la espermatogénesis mediante ubiquitinación (Sutovsky et al. 2000).

Sin embargo, recientemente se ha cuestionado este dogma sobre la herencia del ADNmt a raíz de la aparición de un estudio realizado sobre un paciente con intolerancia severa al ejercicio, causada por una miopatía mitocondrial como consecuencia de una delección de dos bases en el gen *ND2* (Schwartz and Vissing 2002). Esta mutación solamente se detectaba en el tejido muscular del paciente, en una proporción del 90%, pero no en sangre. Además, se determinó que este perfil era de origen paterno, mientras que el ADNmt extraído a partir de otros tejidos coincidía con el perfil materno (Schwartz and Vissing 2002). En base a este hallazgo se ha sugerido que puede haber transferencias del ADNmt paterno en el óvulo fertilizado, de manera que los patrones de herencia podrían ser diferentes a los considerados hasta el momento (Schwartz and Vissing 2003).

Para confirmar o rechazar ese nuevo hallazgo, se llevaron a cabo investigaciones con pacientes con miopatías mitocondriales, pero no se encontró ningún otro caso que se pudiera interpretar como herencia paterna (Bandelt et al. 2005b; Filosto et al. 2003; Schwartz and Vissing 2004; Taylor et al. 2003). Por lo

tanto, parece muy probable que el único caso de herencia paterna del ADNmt detectado en humanos represente, si no un resultado artefactual, sí un fallo en el reconocimiento normal y eliminación de las mitocondrias paternas. De esta forma, la herencia materna del ADNmt aún puede ser considerada como la norma (Taylor et al. 2003).

Otro principio de la biología molecular que también fue cuestionado es la ausencia de recombinación del ADNmt. Entre los años 1999-2000 aparecieron varios trabajos (Awadalla et al. 1999; Eyre-Walker et al. 1999; Hagelberg et al. 1999) en los que se reivindicaban que la elevada frecuencia de homoplasias en el genoma mitocondrial no era debida a mutaciones recurrentes en la filogenia sino a eventos de recombinación. En estos estudios se postulaba que este fenómeno podría ocurrir entre el ADNmt materno y paterno ya que este último sobrevive varias horas antes de ser eliminado (Eyre-Walker et al. 1999) o entre copias de ADNmt presentes en el genoma nuclear (NUMTs) (Awadalla et al. 1999). Posteriormente, se comprobó que había errores en los datos y en los métodos estadísticos empleados en estos trabajos (Elson et al. 2001; Hagelberg et al. 2000; Kivisild et al. 2000; Macaulay et al. 1999).

Una estrategia para evaluar la existencia de recombinación en el ADNmt podría ser el análisis de polimorfismos próximos a posiciones homoplásicas. Sin embargo, no se han encontrado polimorfismos ligados a estas posiciones, indicando que estas mutaciones surgieron de forma aislada (Finnilä et al. 2001).

Dos años después de cuestionar el origen materno del ADNmt (Schwartz and Vissing 2002) se publicó un trabajo que afirmaba que se había observado recombinación entre ADNmt materno y paterno en aproximadamente el 0'7% del ADNmt total en el tejido muscular del mismo paciente (Kraytsberg et al. 2004).

A raíz de todas estas publicaciones, en las que se observaban tipos recombinantes del genoma mitocondrial para un mismo individuo, y que corresponderían con diferentes partes de la filogenia del ADNmt, se llevó a cabo un reanálisis de los mismos. Se concluyó que aparentemente el fenómeno de “mosaicos de ADNmt” podría deberse a contaminación o mezcla entre muestras distintas y no a fenómenos de recombinación (Bandelt et al. 2005b).

Sin embargo, hay autores que aún no descartan la posible existencia de recombinación en el ADNmt humano, y algunos sugieren que el empleo de tests

estadísticos inadecuados pueda ser la causa de no haber podido detectarla (Hagelberg 2003; White and Gemmell 2009).

En resumen, si la fuga de ADNmt paterno es un fenómeno tan poco frecuente como parece, la probabilidad de que pueda ocurrir recombinación, si es que existe, es muy reducida y no afectaría a las interpretaciones obtenidas a partir de los datos del genoma mitocondrial. Por ello, aunque se ha sugerido que es importante tener en cuenta la posibilidad de herencia paterna (Kraytsberg et al. 2004), siempre hay que considerar y descartar antes toda posibilidad de recombinación artefactual (Bandelt et al. 2005b).

## **2.3 Tasa de mutación del genoma mitocondrial**

La tasa de sustitución o evolución depende de la velocidad a la que surgen (tasa de mutación) y se fijan (tasa de fijación) las mutaciones en los linajes. Se ha estimado, mediante estudios comparativos entre las secuencias de ADN de diferentes organismos, que la tasa de sustitución correspondiente al ADNmt es aproximadamente diez veces superior a la del ADNn (Brown et al. 1979), aunque parece que esta diferencia no es uniforme a lo largo de la molécula (Pesole et al. 1999). La ausencia de histonas que protejan el genoma del orgánulo, la fidelidad reducida de los sistemas de replicación y/o reparación del ADNmt y la elevada concentración de especies reactivas de oxígeno (ROS) en la matriz mitocondrial son algunas de las causas que se barajan para explicar la diferencia entre las tasas de mutación de estos dos genomas.

Esta alta tasa de evolución convierte el ADNmt en una herramienta muy útil en genética forense y para el estudio de las poblaciones humanas, ya que permite llevar a cabo análisis de alta resolución de los eventos evolutivos recientes e introducir una escala temporal en la evolución molecular.

La tasa de sustitución no es igual en todas las regiones de la molécula de ADNmt ya que, por ejemplo, la región control se caracteriza por presentar una tasa más elevada que la región codificante, aunque es difícil dar una cifra ya que es un tema muy controvertido. Esta diferencia en tasa de sustitución puede deberse a que en la región control la presión selectiva es mucho menor. De hecho, posiciones



concretas del ADNmt presentan tasas de sustitución muy superiores a la tasa promedio para esta misma zona, llegando algunas a considerarse puntos calientes de mutación o *hotspots* que serían posiciones más propensas a mutar.

Se han llevado a cabo dos tipos de aproximaciones diferentes para estimar la tasa de mutación del ADN mitocondrial. Una de ellas se basa en estudios filogenéticos, en los que se realizan comparaciones inter e intraespecíficas, mientras que la otra estrategia se basa en el estudio de pedigrís o familias que pueden seguirse hacia atrás varias generaciones.

Estimas obtenidas a partir de **estudios familiares** y con datos de la región control aportaron valores bastante diferentes entre sí: Howell et al. (1996) obtienen una tasa de  $7.5 \times 10^{-7}$  mutaciones por sitio y año, Parsons et al. (1997) sitúan esta cifra en  $2.5 \times 10^{-6}$  sustituciones/nucleótido/año, Sigurgardottir et al. (2000) en  $3.2 \times 10^{-7}$  sustituciones/sitio/año, Heyer et al. (2001) en  $3.9 \times 10^{-7}$  sustituciones/sitio/año, Santos et al. (2005) en  $2.4 \times 10^{-7}$  sustituciones/sitio/año.

Comparaciones **filogenéticas** proporcionan estimas considerablemente inferiores, entre  $2.2 \times 10^{-8}$ - $3.5 \times 10^{-7}$  sustituciones/nucleótido/año (Forster et al. 1996; Torroni et al. 1994), a las obtenidas en estudios familiares.

La discrepancia entre las estimas basadas en pedigrís y las filogenéticas puede deberse a que las primeras podrían estar sobreestimando el efecto de heteroplasmías y posiciones *hotspots*, mientras que los estudios filogenéticos tendrían también en cuenta aquellas posiciones que evolucionan más lentamente. Sin embargo, se ha argumentado que no hay un único factor que explique esta discrepancia, sino que sería el resultado combinado de la influencia de *hotspots* mutacionales, deriva génica, selección purificante y mutaciones recurrentes (Howell et al. 2003; Sigurgardóttir et al. 2000).

Se ha propuesto que la aproximación más adecuada para estudiar eventos de historia antigua sería la basada en estudios filogenéticos, porque representan mutaciones que han alcanzado una frecuencia considerable en la población, mientras que para eventos recientes podría ser más apropiado emplear estimas obtenidas a partir de estudios basados en pedigrís, aunque las primeras también funcionan bien para eventos de 1000 años (Macaulay et al. 1997). Ingman (2000) empleó una tasa de sustitución de  $1.7 \times 10^{-8}$  sustituciones/sitio/año

(excluyendo la región control) para datar el origen del hombre moderno. La tasa de mutación del ADNmt empleada con más frecuencia es  $1'26 \times 10^{-8}$  sustituciones/nucleótido/año (Mishmar et al. 2003), pero sólo tiene en cuenta la variabilidad de la región codificante y no considera el efecto de la selección purificante. La observación de que mutaciones no sinónimas son más frecuentes en ramas jóvenes de la filogenia llevó a pensar que podría deberse a la selección purificante. Así, se propuso una nueva calibración basada en la tasa de acumulación de transiciones sinónimas en el genoma mitocondrial, aunque sólo emplea una pequeña fracción del mismo para su cálculo ( $3'5 \times 10^{-8}$  transiciones sinónimas/sitio/año) (Kivisild et al. 2006).

Sin embargo, las estimas promedio de la tasa de sustitución no reflejan el estado real de los hechos, ya que las tasas varían entre los genes codificantes, los de ARNt o ARNr, dentro de la región control, y entre posiciones no sinónimas y sinónimas (Pesole et al. 1999). Aunque es muy importante estimar la tasa de sustitución promedio del ADNmt, desde otros puntos de vista (e.j. forense), es mucho más relevante tener información sobre las tasas de mutación sitio-específicas.

Recientemente, se vio la necesidad de devolverle al ADNmt la credibilidad en lo que a la genética evolutiva se refiere, proponiendo una calibración que corrija la tasa de mutación en función de su dependencia temporal (Soares et al. 2009). Este trabajo confirma el efecto de la selección purificante, que actúa sobre mutaciones ligeramente deletéreas, en ramas jóvenes del árbol mitocondrial. Por ello, se propone un reloj molecular mejorado en el que se considera la variabilidad de todo el genoma mitocondrial y el efecto de selección purificante, obteniendo una tasa de sustitución para la molécula completa de  $1'665 \times 10^{-8}$  sustituciones/nucleótido/año o una mutación cada 3624 años. Este estudio aporta estimas para la tasa de sustitución de la región codificante, la región control o la tasa de sustituciones sinónimas, además de estimas de la edad de las principales ramas del árbol mitocondrial. Se indican también las posiciones *hotspots* más importantes, lo que puede ser de gran ayuda en el campo forense. Con esta nueva aproximación se obtienen estimas que coinciden con dataciones obtenidas a partir de registros

arqueológicos, proporcionando un acercamiento entre las evidencias fósiles y las genéticas.

## 2.4 Variabilidad del ADN mitocondrial

El término polimorfismo fue descrito por Ford en 1940 como la "presencia en una población de dos o más formas alélicas discretas, de las que la de menor frecuencia no puede mantenerse sólo por mutación recurrente". En términos generales se asume que un locus es polimórfico cuando el alelo más común tiene una frecuencia inferior al 99%.

El ADN codificante es, en general, poco variable o polimórfico entre los individuos, mientras que las regiones no codificantes, al no estar sujetas a fuerte presión selectiva, presentan mucha mayor variabilidad. Esta es una de las razones que promovió la sustitución del análisis de los polimorfismos clásicos (proteicos) por los polimorfismos de ADN de regiones no codificantes. Los polimorfismos están distribuidos a lo largo de todo el genoma, y se pueden clasificar en polimorfismos de longitud o de secuencia.

Los **polimorfismos de longitud** son causados por inserciones o deleciones de uno o más nucleótidos. Este tipo de polimorfismos son los más frecuentes en el ADN repetitivo (minisatélite y microsatélite), en donde las variaciones entre los alelos se deben a diferencias en el número de copias de la unidad repetitiva. En el ADNmt existen trectos homopoliméricos en los que con cierta frecuencia ocurren inserciones o deleciones de uno o varios nucleótidos. Estos trectos están localizados en las posiciones 16184-16188/16190-16193, 303-309/311-315 y 568-573 de la región control. La explicación más probable para este tipo de polimorfismos es el deslizamiento de la polimerasa durante la replicación del ADN, y este efecto parece ser más frecuente cuando el número de unidades repetitivas sobrepasa un valor umbral. Se ha sugerido que el número crítico para el tracto de poli-citosinas de HVI son 8Cs, mientras que para HVII serían 7Cs (Irwin et al. 2009). En el ADNmt sólo se ha descrito un microsatélite corto entre las posiciones 514 y 523 de la región *D-loop* del genoma mitocondrial (Bodenteich et al. 1992).

Los **polimorfismos de secuencia** son causados por el cambio de uno o más nucleótidos en una secuencia de ADN. Cuando el cambio sucede en un único

nucleótido y las dos bases alternativas aparecen en la población en una frecuencia superior a 1%, este tipo de polimorfismo se denomina SNP (*Single Nucleotide Polymorphism*). Los SNPs son el tipo de polimorfismo más abundante del genoma humano y la mayoría se encuentran en regiones donde no alteran la función génica. Debido a su abundancia, su simplicidad y su baja tasa de mutación, estos marcadores han despertado un gran interés en distintos campos como el de la genética forense, el antropológico, y por supuesto también el clínico (Sobrino et al. 2005). En el ADNmt las transiciones superan en número a las transversiones en una proporción aproximada de 15:1, y dentro de las primeras, los cambios nucleotídicos entre pirimidinas son más frecuentes que entre purinas (Irwin et al. 2009).

## 2.5 Número de copias y heteroplasmía

Cada célula contiene entre cientos y miles de mitocondrias y a su vez cada mitocondria contiene múltiples copias de la molécula de ADN del orgánulo, al contrario que en el ADN nuclear, donde generalmente sólo existen dos copias de un gen o segmento de ADN. Esta propiedad hace del ADNmt la molécula de elección para el análisis de ADN antiguo y para ciertas aplicaciones forenses donde la obtención de ADNn en buenas condiciones es limitante.

Normalmente, las moléculas de ADNmt se distribuyen por la matriz mitocondrial formando complejos nucleoproteicos, llamados nucleoides, que representan las unidades dinámicas y de herencia. Se estima que hay cientos de nucleoides por célula y que cada uno suele contener entre 2-15 moléculas de ADNmt, y parecen estar están unidos a la membrana interna (Malka et al. 2006). Se desconoce cómo está empaquetado el ADNmt, pero se sabe que carece de la protección de histonas característica del genoma nuclear.

En principio, todas las moléculas de ADNmt de un tejido son idénticas, denominándose a este estado **homoplasmía**, pero en ocasiones dentro de una mitocondria, de una célula, en un tejido específico o dentro de un individuo puede coexistir más de un tipo de ADNmt. La existencia de diferentes moléculas de ADNmt dentro de un individuo se conoce como **heteroplasmía**.

La región control del genoma mitocondrial, muy interesante desde el punto de vista forense y antropológico por ser muy polimórfica (Greenberg et al. 1983), es el área más frecuentemente analizada y por ello la mayoría de las estimas de tasas de heteroplasmía se basan en estos datos. Estimaciones iniciales atribuían un valor pequeño o nulo a este fenómeno (Monnat and Reay 1986), sin embargo, estudios posteriores reportaron tasas de heteroplasmías puntuales considerablemente superiores: 13'8% para la región HVI (Tully et al. 2000), 11'6% en la región HVII (Calloway et al. 2000) y 10'5% para ambas regiones (Paneto et al. 2007).

Las heteroplasmías de longitud se detectan frecuentemente en los tractos homopoliméricos de HVI y HVII, mientras que el STR dinucleotídico (AC)<sub>n</sub>, presenta un bajo grado de heterocigosidad. La utilidad forense de este microsatélite es reducida por su herencia inestable, ya que las heteroplasmías pueden aparecer y desaparecer en pocas generaciones (Szibor et al. 2007). En general, las heteroplasmías de longitud se detectan con más frecuencia (52%) que las heteroplasmías puntuales (6%), que habitualmente coinciden con las posiciones nucleotídicas más hipervariables (Irwin et al. 2009).

Es importante, además, tener en cuenta que no todas las metodologías empleadas para detectar heteroplasmías presentan la misma sensibilidad. Por ejemplo, la secuenciación directa permite detectar heteroplasmías en las que la proporción del alelo minoritario se encuentre entre 10-20% (Bendall et al. 1996; Irwin et al. 2009), mientras que la pirosecuenciación presenta mucha mayor sensibilidad (1%) (White et al. 2005).

Aunque pueden existir posiciones heteroplásmicas en cualquier muestra biológica, algunos tejidos son más susceptibles a sufrir mutaciones y por tanto a presentar dos variantes alélicas para una misma posición, ya sea por sus mayores requerimientos energéticos o por el efecto de cuellos de botella más marcados. Calloway et al. (2000) encontraron que el músculo esquelético era el tejido con mayor tasa de heteroplasmías, en comparación con tejido cardíaco, cerebral y sanguíneo, incrementando su frecuencia con la edad del individuo. También el pelo parece presentar tasas de heteroplasmía superiores al tejido sanguíneo, y las proporciones de las variantes alélicas pueden variar considerablemente entre

distintas secciones de un mismo pelo (Paneto et al. 2007; Salas et al. 2001a; Tully et al. 2004).

Puede que el concepto de homoplasmía sea más aparente que real ya que teniendo en cuenta el alto número de copias del ADNmt y a la elevada tasa de mutación de este genoma, es poco probable la existencia de un único tipo de ADNmt en un individuo. Sin embargo, recientemente han surgido estudios que sugieren que determinados artefactos, como mutaciones fantasma (Brandstätter et al. 2005) y detección de NUMTs (Parr et al. 2006), podrían provocar un porcentaje de heteroplasmías en los electroferogramas superior a la tasa real.

### 3 Análisis de la variabilidad del ADN mitocondrial

De acuerdo con lo mencionado anteriormente, las diferencias entre secuencias de ADNmt son debidas únicamente a mutación, ya que este genoma no experimenta recombinación. El análisis de la variabilidad de este pequeño genoma ha despertado el interés de tres áreas de la genética: poblacional, forense y clínica.

#### 3.1 Evolución de la metodología de genotipado del ADNmt

La primera aproximación para el análisis de la variabilidad del ADNmt consistió en la digestión del ADN mediante un enzima de restricción (Denaro et al. 1981) y posterior separación electroforética (**RFLPs** o *Restriction Fragment Length Polymorphisms*). Esta metodología se utilizaba fundamentalmente con fines poblacionales, ya que se requería gran cantidad de ADN no degradado.

A finales de la década de los años 80 y durante la de los 90, surgieron otras tecnologías para analizar la variabilidad del ADNmt como la **SSCP** (*Single Strand Conformation Polymorphism*) (Orita et al. 1989). Esta técnica se basa en la diferente movilidad de las hebras de ADN de una muestra previamente desnaturalizada sobre un gel de acrilamida, en función de la conformación que adopten dichas hebras durante el recorrido electroforético. Es una técnica de *screening* sencilla, barata y rápida, que se utilizó para la detección de mutaciones puntuales, pero presentaba una difícil estandarización entre laboratorios y sólo detectaba un 90% de las mutaciones. Posteriormente, se desarrollaron variaciones de esta técnica,

como la **RE-SSCP** (*Restriction Enzyme-SSCP*)(Barros et al. 1997) o la **FSSCP-OF** (*Fluorescent-SSCP-Overlapping fragments*) (Salas et al. 2001b).

El desarrollo de la técnica de la **PCR** (*Polymerase Chain Reaction*) por Kari Mullis (1986) supuso un gran avance en el análisis de la variación genética, ya que permite obtener un gran número de copias de un fragmento de ADN particular a partir de una cantidad reducida de muestra. Como consecuencia, se adaptaron las tecnologías de análisis de la variabilidad del ADNmt, surgiendo así, por ejemplo, la metodología PCR-RFLP (Torroni et al. 1992).

El **CSGE** (*Conformation-Sensitive Gel Electrophoresis*) es una técnica, basada en un principio similar al SSCP, que permite hacer un *screening* para detectar polimorfismos o mutaciones (Finnilä et al. 2000). Consiste en la amplificación del fragmento de ADN de interés, la adición del fragmento correspondiente de una muestra de secuencia conocida, desnaturalización para permitir la formación de heterodúplex y separación mediante electroforesis. Estos heterodúplex migrarán de forma diferente en el gel de acrilamida si existen mutaciones, lo que permitirá determinar qué fragmentos se deben secuenciar.

Sin embargo, el método de elección para la mayoría de los laboratorios fue la **secuenciación**, ya que permite conocer exactamente qué mutación está presente en el ADN. En el año 1977, se publicaron dos métodos de secuenciación distintos: secuenciación química (Maxam and Gilbert 1977) y secuenciación enzimática (Sanger et al. 1977). La **secuenciación química** se basa en la ruptura en bases específicas (A, T, G o C) de una molécula de ADN marcada en su extremo 5' para generar una serie de fragmentos de diferentes longitudes que posteriormente se separan mediante electroforesis en gel. El otro método alternativo es la **secuenciación enzimática**, en la que se generan fragmentos de diferentes tamaños debido a la interrupción de la síntesis de una cadena de ADN en bases específicas, como consecuencia de la incorporación de análogos de nucleótidos.

En la actualidad se utiliza casi universalmente el método dideoxy o enzimático, basado en la metodología de secuenciación de Sanger, puesto que puede aplicarse a fragmentos más grandes, es más rápida y permite ser automatizada. Este sistema requiere el uso de 2'-3'-dideoxinucleótidos trifosfato (ddNTPs), que difieren de los deoxinucleótidos (dNTPs) en la sustitución del grupo

3'-hidroxilo por un hidrógeno; además, cada uno de los cuatro ddNTPs posee marcajes fluorescentes diferentes. Los ddNTPs se añaden al extremo 3' de la cadena creciente, pero al carecer del grupo 3'-hidroxilo no puede unirse el siguiente nucleósido 5'-trifosfato entrante, por lo que se detiene la síntesis. De esta manera, la inclusión de niveles bajos de dideoxinucleótidos en el medio permite que la síntesis de ADN termine al azar en cualquier sitio, obteniéndose cientos de fragmentos de longitudes que difieren en un único nucleótido.

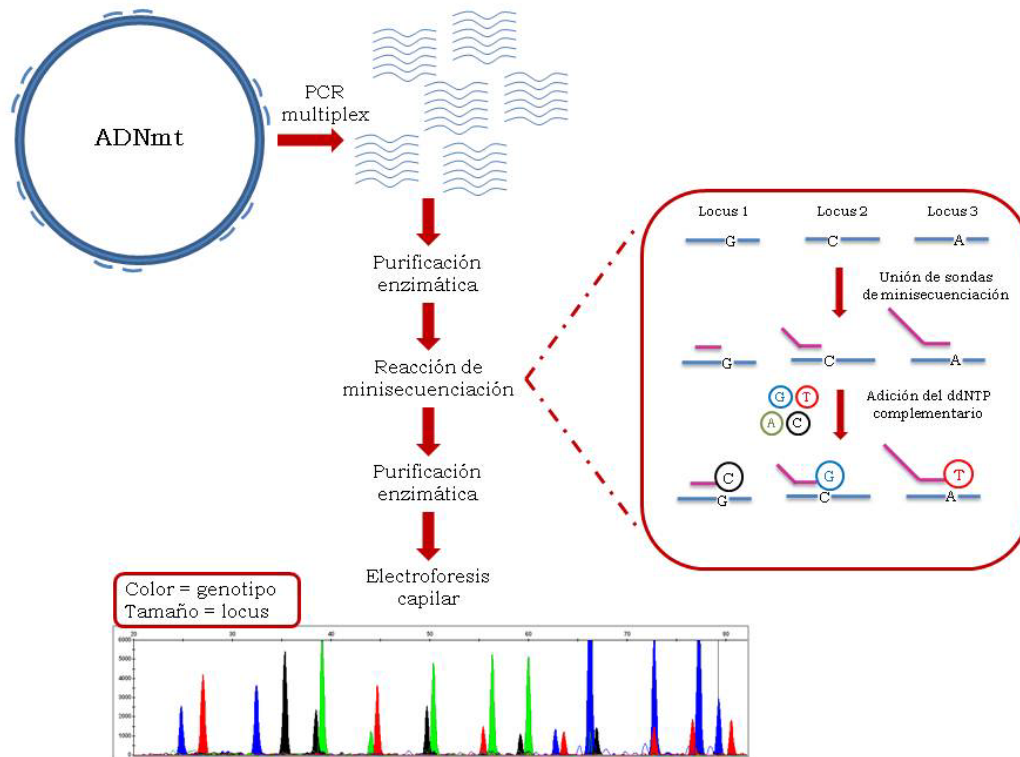
El sistema que se utiliza actualmente de forma mayoritaria es la secuenciación cíclica, que combina la técnica de la PCR con la secuenciación enzimática, y que incrementa la sensibilidad del método (Sullivan et al. 1991).

Hasta hace relativamente poco tiempo, la región de elección para el análisis del ADNmt era el segmento hipervariable I (HVI) de la región control, definida por Vigilant (1989) y comprendida entre la posición 16024 y 16365, ya que era la región que presentaba un grado más alto de variación de secuencia (Sullivan et al. 1991). Posteriormente, se incorporó a la rutina el análisis de la región hipervariable II (HVII) situada entre las posiciones 73 y 340, de manera que pasaban a evaluarse alrededor de 610pb, aumentando así el poder de discriminación. Sin embargo, la tendencia es poder analizar cada vez más posiciones, y así en estos últimos años la recomendación es secuenciar la región control completa, comprendida entre la posición 16024 hasta la 16569 y desde la 1 hasta la 576. Sin embargo, a pesar de que la región control del ADNmt es el área con mayor variabilidad de secuencia, tiene un poder de discriminación limitado, no sólo en el contexto forense, sino también desde un punto de vista antropológico o clínico. Esto depende en gran medida de la población de estudio, ya que la filogenia varía enormemente entre las distintas poblaciones, pero algunos haplogrupos están muy mal definidos en las regiones hipervariables, de manera que perfiles que coincidan en estas regiones no tienen por qué pertenecer necesariamente al mismo haplogrupo ni por supuesto al mismo individuo (Álvarez-Iglesias et al. 2009; Quintáns et al. 2004). Además, la región control presenta la tasa de mutación más alta del genoma mitocondrial, por lo que son frecuentes las posiciones recurrentes en la filogenia (homoplasia), pudiendo llevar a errores de interpretación (Chen et al. 1995).



La situación ideal, sería poder secuenciar siempre genomas completos, sin embargo esto es muy laborioso, costoso y requiere gran cantidad de muestra, por lo que resulta totalmente inviable en la rutina forense y poco frecuente en otros contextos. El análisis de SNPs de la región codificante del genoma mitocondrial ha resultado ser una buena estrategia para solventar estos problemas. El creciente interés por estos marcadores en los últimos años ha ido paralelo al espectacular desarrollo de nuevas tecnologías para su genotipado, que permiten el análisis de varios marcadores a la vez mediante multiplex. El abanico de tecnologías para el genotipado de SNPs actualmente es enorme, sin embargo, a la hora de seleccionar la metodología más adecuada deben tenerse en cuenta aspectos como la sensibilidad, reproducibilidad, precisión, capacidad para analizar reacciones multiplex, volumen de SNPs, calidad/cantidad de las muestras, coste y flexibilidad de la tecnología (Sobrino et al. 2005).

En el caso del ADNmt el método más popular para el análisis de la variabilidad de la región codificante es la técnica de **minisecuenciación**, concretamente empleando el kit de SNaPshot™ (Applied Biosystems). La estrategia de minisecuenciación o SBE (*Single Base Extension*) es un sistema de detección de SNPs basado en la adición de un único dideoxinucleótido (ddNTP), marcado fluorescentemente, al extremo 3' de un *primer* específico y adyacente al polimorfismo. La reacción multiplex puede llevarse a cabo porque los productos de minisecuenciación llevan incorporadas colas no homólogas en su extremo 5' que permiten que puedan ser separados por tamaño mediante electroforesis en los secuenciadores automáticos.



**Figura 5:** Representación gráfica de la técnica de minisequenciación para el genotipado de SNPs del ADN mitocondrial.

Esta técnica fue descrita por Tully et al. (1996) para genotipar posiciones de la región control, sin embargo, el grueso de trabajos realizados con esta técnica ha aparecido en los últimos años y en ellos se analizan también SNPs de la región codificante (Álvarez-Iglesias et al. 2008; Álvarez-Iglesias et al. 2009; Brandstätter et al. 2003; Quintáns et al. 2004; Vallone et al. 2004). El éxito de esta técnica probablemente resida en su capacidad para interrogar un número considerable de SNPs en una única reacción y en su bajo coste.

La **secuenciación de genomas completos** es la única metodología que permite analizar toda la variabilidad del genoma mitocondrial, sin embargo no se emplea de forma generalizada por ser muy laboriosa y de coste elevado. Aún así, en determinados estudios de genética clínica y genética de poblaciones, la secuenciación de genomas completos es la tendencia actual. En un futuro cercano, gracias al desarrollo de tecnologías de ultrasecuenciación, el tiempo requerido para secuenciar un genoma completo disminuirá en gran medida y puede que esto mejore la accesibilidad a esta estrategia.

### 3.2 Nomenclatura

El estándar universal para referenciar la variabilidad del ADN mitocondrial, deriva de la primera secuencia completa publicada de un genoma mitocondrial, denominada “*Cambridge Reference Sequence*” (CRS) o también secuencia de Anderson (Anderson et al. 1981). Cada una de las bases del genoma mitocondrial se nombra con un número, en función de la posición que ocupa en el genoma, y tomando como punto de inicio una posición cercana al origen de replicación de la cadena H, siendo esta base la posición 1, y la anterior la base 16569. Posteriormente, se comprobó que la secuencia de Anderson contenía algunos errores, sobre todo en la región codificante, de ahí que en la actualidad se utilice la versión corregida de la misma (“*Revised Cambridge Reference Sequence*” o rCRS) como secuencia de referencia (Andrews et al. 1999).

A la hora de analizar una muestra, se anotan únicamente aquellas posiciones que difieren de la secuencia de referencia, quedando implícito que el resto de la secuencia es exactamente igual a la de Anderson. Se acordó referenciar estas diferencias con respecto a la cadena L, pero no existe tanto consenso sobre cómo expresar las variantes de dicha secuencia.

En los últimos años, gracias a grupos como la ISFG (*International Society of Forensic Genetics*) o al grupo EDNAP (*European DNA Profiling Group*) parece que se ha llegado a cierto acuerdo en lo relativo al uso del ADNmt en el campo forense. Se han desarrollado guías (Tully et al. 2001) que proponen, entre otras cosas, una nomenclatura estándar para expresar las variantes del genoma mitocondrial:

- **Cambio de base:** se indica la posición en la cual se ha detectado una variante con respecto a rCRS y tras ella la letra correspondiente a la nueva base. (Ejemplo: Si en la posición 16519 en rCRS existe una citosina (C) y en la secuencia de estudio hay una timina (T), se designará como: “16519T”).
- **Deleciones:** se informa la posición delecionada. (Ejemplo: si la base 249 ha sido delecionada, debe referenciarse como 249d).
- **Inserciones:** deben ser designadas con “.1” después de la posición de menor número entre las dos bases donde ha ocurrido la inserción. (Ejemplo: si entre las posiciones 245 y 246 ha tenido lugar la inserción de una adenina (A), se

designará como “245.1A”. En los tractos homopoliméricos no es posible saber en qué posición ha tenido lugar la inserción, y en estos casos se asume que ha ocurrido al final del tracto, en la base con mayor numeración. (Ejemplo: Si la inserción ha tenido lugar en el tracto poli-C situado entre las posiciones 302 y 310, se denotará como “309.1C”, y si existen dos inserciones “309.1C y 309.2C”).

- **Heteroplasmías:** las heteroplasmías son el resultado de detectar la existencia de dos tipos de moléculas en una misma muestra. Siempre deben secuenciarse en ambas direcciones para su confirmación, y si no hay acuerdo, se designan como ambiguo (N).
  - *Heteroplasmías puntuales:* los tipos de moléculas difieren en una sola base. Si la proporción de cada tipo es similar, la designación puede hacerse empleando el código IUB/IUPAC (Tabla 3) o ambas variantes. (Ejemplo: 152Y o 152T~C). Sin embargo, si la proporción de una de las dos clases de moléculas es mayoritaria, debe denotarse como “C>T”.
  - *Heteroplasmías de longitud:* la diferencia entre los tipos moleculares consiste en una inserción o deleción de una o más bases en alguno de ellos. Son más difíciles de confirmar porque desfasan la pauta de lectura. (Ejemplo: 309.1C>309.2C)

Código	Significado	Código	Significado
A	Adenina	S	C/G
C	Citosina	Y	C/T
G	Guanina	K	G/T
T	Timina	V	A/C/G
U	Uracilo	H	A/C/T
M	A/C	D	A/G/T
R	A/G	B	C/G/T
W	A/T	N	A/G/C/T

**Tabla 3:** Código IUB/IUPAC, empleado para representar secuencias de ácidos nucleicos.

Algunos autores (Salas et al. 2005a) propusieron una nomenclatura alternativa, más compacta, según la cual sólo se haría referencia a la base mutada cuando se trate de una transversión (Ejemplo: 16183C, 16189), y las inserciones se reportarían en su conjunto (Ejemplos: 523+ACAC, 309+CC). Según la nomenclatura

propuesta para los estudios de genética clínica, una mutación se debería referenciar de la siguiente forma: m.3243A>G. La “m” indica que se trata de una mutación en el ADNmt, y el símbolo “>” indicaría el sentido del cambio de base.

A pesar de la gran utilidad de estas guías, hay situaciones especialmente complejas en las que no está claro cuál es la forma más correcta de alinear dos secuencias. Para la resolución de estos casos se han propuesto unas sencillas reglas a la hora de registrar las variantes con respecto a la rCRS (Wilson et al. 2002):

- Caracterizar perfiles usando el menor número de diferencias posibles.
- Si hay más de una forma para mantener el mismo número de diferencias, éstas deben ser priorizadas de la siguiente forma:
  - Inserciones/deleciones
  - Transiciones
  - Transversiones
- Las inserciones y deleciones deben estar situadas en posición 3’ con respecto a la cadena L, y mejor de forma contigua que separadamente. Cuando hay que combinar inserciones/deleciones y sustituciones, es recomendable que el cambio de base esté situado en posición más 5’ mientras que la inserción/delección en posición 3’.

Sin embargo, en algunas ocasiones, especialmente en eventos de *indels* (inserciones-deleciones), seguir estas reglas va en contra del mecanismo más probable por el que surgieron las variantes. En estos casos la información filogenética puede aportar pistas sobre los alineamientos alternativos más naturales, a pesar de que pueden suponer un mayor número de diferencias. En el campo forense, como consecuencia de un mal alineamiento, podría asignarse erróneamente el número de inconsistencias entre dos muestras y esto afectaría a su interpretación (Bandelt and Parson 2008).

## 4 Aplicaciones poblacionales de la variabilidad del genoma mitocondrial

### 4.1 Generalidades

El análisis del genoma humano ha permitido conocer diversos aspectos sobre el origen y la evolución de nuestra especie. Inicialmente se creía que individuos pertenecientes a distintas regiones geográficas o a diferentes grupos étnicos debían presentar numerosas diferencias en su genoma. Sin embargo, sólo un 3-5% de la variación genética se debe al componente interpoblacional, mientras que un 93-95% es explicado por el componente intrapoblacional (Rosenberg et al. 2002). STRs y SNPs autosómicos, presentan gran utilidad en el campo de genética de poblaciones porque pueden, por ejemplo, aportar información sobre la ancestralidad de una muestra (Phillips et al. 2007), aunque su análisis es mucho más complejo que el de los marcadores uniparentales, cromosoma Y (región no recombinante) y ADNmt. Las filogenias del genoma mitocondrial, heredado por vía materna, y de la porción no recombinante del cromosoma Y, heredado por vía paterna, mantienen registros secuenciales de la acumulación de la diversidad genética humana. Es precisamente esta característica la que los ha convertido en las herramientas de elección para la genética de poblaciones, a pesar de que la información contenida en el genoma humano entero será necesaria para caracterizar completamente la evolución genética humana (Underhill and Kivisild 2007).

Las características únicas del ADN mitocondrial humano (elevado número de copias, herencia materna, ausencia de recombinación y alta tasa de mutación) hacen que esta molécula sea esencial en el campo de la genética de poblaciones, como demuestran muchos estudios relacionados con la historia y evolución de las poblaciones humanas. Gracias al análisis de esta pequeña molécula han surgido hipótesis de gran relevancia en relación al origen del hombre y su dispersión a lo largo del planeta.

La filogenia molecular establece relaciones evolutivas (dentro de una misma población, entre poblaciones de una misma especie o entre especies) basándose en

datos moleculares. Mediante el análisis de la variación de secuencia, muestreada en el presente, y basándose en el principio de coalescencia, se pueden inferir las relaciones filogenéticas entre secuencias individuales (haplotipos) y el ancestro materno común a todas ellas (MRCA, *Most Recent Common Ancestor*). Existen diferentes procedimientos para la reconstrucción filogenética, pero el más utilizado es el de “máxima parsimonia” o de “evolución mínima”, y las representaciones gráficas de las relaciones filogenéticas son los árboles filogenéticos. La estructura de este árbol genealógico contiene información demográfica que, en conjunto con una tasa de mutación adecuada, permite estimar la escala temporal para eventos de la historia del hombre moderno.

A lo largo de la historia del hombre moderno, se han ido acumulando secuencialmente mutaciones en la molécula de ADNmt, de forma que las secuencias fueron divergiendo unas de otras dando lugar a linajes independientes que se conocen como haplotipos. Los haplotipos mitocondriales pueden ser agrupados en haplogrupos, conjuntos de secuencias relacionadas por descendencia y que pertenecen a una misma rama monofilética del árbol filogenético. Los haplogrupos están definidos por motivos, es decir, uno o varios polimorfismos puntuales, estables y antiguos que son compartidos por todas las secuencias de esa rama (Richards et al. 1998; Salas et al. 2007).

El término “filogeografía” es descrito en 1987 por Avise y hace referencia al estudio de la distribución geográfica de los distintos linajes o haplogrupos mitocondriales. El análisis de la variabilidad mitocondrial bajo esta perspectiva ha revelado la asociación entre determinados haplogrupos y regiones geográficas o grupos étnicos concretos, aunque por sí solo no aporta una evidencia total sobre el origen geográfico de una muestra. Esta estrategia ha permitido deducir movimientos poblacionales y reconstruir la historia evolutiva de nuestra especie (Richards et al. 1998).

## 4.2 Metodología de análisis

Las posiciones diagnóstico de linajes mitocondriales se encuentran dispersas en todo el ADNmt, pero en la región control, a causa de su mayor tasa de mutación

y/o menor presión selectiva, pueden aparecer más de una vez en la filogenia (homoplasia). Por ello, una clasificación basada en un bajo número de polimorfismos, especialmente si se encuentran en la región control, puede oscurecer y dificultar la interpretación de la filogenia, de forma que es recomendable considerar la variabilidad del ADNmt en conjunto (Chen et al. 1995; Finnilä et al. 2001; Salas et al. 2007).

Hasta hace relativamente poco tiempo, la mayor parte de los estudios de la variación del ADNmt en las distintas poblaciones humanas se habían realizado principalmente por dos métodos: secuenciación directa de la región control (Salas et al. 1998; Vigilant et al. 1991) o por amplificación de la molécula completa de ADNmt en fragmentos solapantes y posterior digestión con enzimas de restricción (Chen et al. 1995; Torroni et al. 1992).

En los últimos años, han surgido nuevas tecnologías de alto rendimiento, como CSGE seguido de secuenciación (Finnilä et al. 2001) o secuenciación de genomas completos, que permiten explorar un mayor porcentaje o la totalidad de la variabilidad del genoma mitocondrial, respectivamente. Esta última estrategia es muy laboriosa y supone un gran esfuerzo económico, sin embargo desde sus comienzos (Maca-Meyer et al. 2001; Rieder et al. 1998; Torroni et al. 2001), el número de publicaciones en las que se ha empleado esta metodología ha aumentado considerablemente (Achilli et al. 2008; Álvarez-Iglesias et al. 2009; Herrnstadt et al. 2002; Kong et al. 2006; Yao et al. 2006). Teniendo en cuenta el esfuerzo que supone trabajar con genomas completos, resulta una buena estrategia combinar los datos de la región control con posiciones concretas de la región codificante, analizadas tradicionalmente con RFLPs y en los últimos años con otras tecnologías como la minisecuenciación con SNaPshot™ (AB) (Álvarez-Iglesias et al. 2007; Álvarez-Iglesias et al. 2009; Brandstätter et al. 2003; Quintáns et al. 2004; Roostalu et al. 2007).

#### **4.3 Principales aportaciones del ADNmt en genética de poblaciones**

A partir del estudio de la variación genética del genoma mitocondrial se ha avanzado de forma considerable en el conocimiento de eventos tan importantes



como el origen del hombre moderno y los movimientos migratorios que dieron lugar a la colonización de los distintos continentes. También este genoma ha permitido datar estos y otros acontecimientos gracias al empleo de estimas de la tasa de mutación de este marcador.

#### 4.3.1 Origen del hombre moderno

En 1981 se publica el primer trabajo (Denaro et al. 1981) en el que se trata de establecer las relaciones filogenéticas entre las poblaciones “africana”, “oriental” y “caucásica” estudiando la variabilidad del ADNmt mediante el uso de un enzima de restricción. Encontraron diferencias en las frecuencias de los fragmentos originados entre los distintos grupos étnicos y sugirieron que todos derivaban de un origen asiático.

Rebecca Cann et al., en 1987, siguiendo una línea similar al trabajo anterior, digieren el ADNmt de individuos de poblaciones de Asia, Australia, Nueva Guinea, Europa y África con 12 enzimas de restricción, lo que le permite analizar un mayor porcentaje de la variabilidad de este genoma (Cann et al. 1987). En base a los resultados obtenidos, formulan una hipótesis (bautizada como la hipótesis de la “*Eva mitocondrial*” en el argot periodístico) que propone que todos los tipos de ADNmt actuales se remontan a un único ancestro que vivió en África hace aproximadamente 200000 años. La hipótesis formulada por Cann et al. está basada en el principio de coalescencia según el cual, asumiendo que hubo un origen único para todos los organismos vivos, se deduce que toda la variación en cualquier segmento del ADN mitocondrial (o nuclear) en las generaciones presentes deben proceder en última instancia de un único ancestro que existió en generaciones previas. El árbol filogenético resultante de este trabajo tiene una especial relevancia ya que sitúa por primera vez los ADNmt africanos en la raíz del árbol y posteriormente se separan las diferentes ramas no africanas.

Aunque actualmente esta hipótesis es la aceptada mayoritariamente, existió mucho debate entre estas dos corrientes contrapuestas:

- **Hipótesis multiregional:** propone que, después de la salida de África y posterior dispersión del *Homo erectus*, características anatómicamente modernas fueron surgiendo en diferentes regiones y se difundieron entre las

poblaciones gracias al flujo génico entre ellas. Esto explicaría el patrón de evolución compartido entre los humanos modernos y a su vez las diferencias regionales. Esta hipótesis rechaza totalmente la sustitución total de las formas primitivas por las modernas y está apoyada por evidencias fósiles de continuidad cultural y morfológica entre humanos primitivos y modernos fuera de África (Templeton 1997; Wolpoff et al. 2000).

- **Hipótesis del origen africano reciente o “*Out of Africa*”:** postula que el hombre moderno se originó en África hace 100000-200000 años y que posteriormente se dispersó por el resto del Viejo Mundo reemplazando las formas humanas primitivas, con poca o ninguna mezcla genética. Esta interpretación ha sido tremendamente controvertida, fundamentalmente debido a la utilización de determinados métodos de reconstrucción filogenética, pero el volumen de evidencias genéticas a favor de esta hipótesis hace que sea la más aceptada.

Poco más tarde Vigilant et al. (1991) secuenciaron la región control de 189 individuos de diverso origen geográfico y de nuevo en este trabajo se separan las ramas africanas de las no africanas, situándose las primeras en la base del árbol filogenético.

Estudios de PCR-RFLP (Chen et al. 1995; Torroni et al. 1992) revelaron que la mayoría de los ADNmt africanos, asiáticos, europeos y americanos están definidos por al menos un polimorfismo específico de continente. Su alta frecuencia dentro de los principales grupos continentales permite inferir el origen étnico y geográfico, y sugiere que aparecieron posteriormente a la separación de las poblaciones ancestrales humanas que dieron lugar a los principales grupos étnicos modernos. Estudios en los que se analiza el genoma mitocondrial completo (Finnilä et al. 2001; Ingman et al. 2000; Maca-Meyer et al. 2001), apoyan también la hipótesis *Out of Africa* ya que encuentran que la población africana presenta una mayor diversidad de secuencia como resultado de un tamaño poblacional elevado y constante o de una historia genética más larga. Por el contrario, poblaciones no africanas presentan una filogenia “tipo estrella”, lo que sugiere que sufrieron un cuello de botella a partir del cual recientemente experimentaron una expansión poblacional.

Hoy en día se acepta que, como consecuencia de las mutaciones que han ido emergiendo en las secuencias a medida que el hombre moderno colonizaba el planeta, los haplogrupos presentan una especificidad geográfica considerable y se pueden asociar con los principales grupos étnicos. De esta forma podemos decir que los haplogrupos L1, L2 y L3 agrupan los linajes de África sub-Sahariana; H, I, J, K, T, U, V, W, y X, engloban la mayoría de los haplogrupos de las poblaciones de Europa, Norte de África y oeste de Asia central y A, B, C, D, E, F, G y M son los haplogrupos que engloban la mayor parte de los linajes de Asia, Oceanía y nativo americanos.

Las estimas obtenidas en relación a la edad del antepasado común más reciente o TMRCA (*Time to the Most Recent Common Ancestor*) para los humanos modernos varían considerablemente entre los diferentes estudios, ya que las tasas de sustitución empleadas también son distintas y fueron mejorando progresivamente:  $171500 \pm 50000$  (Ingman et al. 2000),  $198000 \pm 19000$  (Mishmar et al. 2003),  $204700 \pm 22100$  (Macaulay et al. 2005),  $192400 \pm 41000$  (Soares et al. 2009).

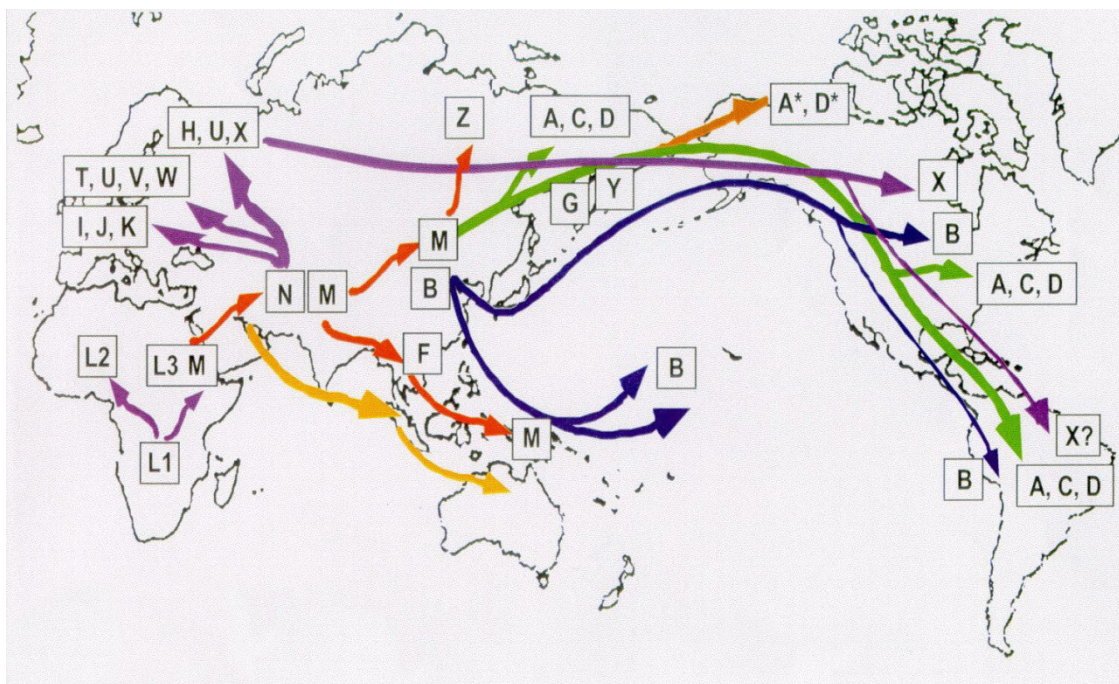
#### 4.3.2 Rutas de dispersión y colonización del planeta

El análisis de la variación del genoma mitocondrial también ha contribuido al conocimiento de las rutas de dispersión de las poblaciones humanas. Existen tres hipótesis principales y con el paso del tiempo el peso que ha recibido cada una de ellas ha ido variando, como consecuencia de la aparición de nuevos datos, tanto genéticos como arqueológicos:

- **Ruta Norte:** esta ruta consistiría en la dispersión por el norte de África y Levante, hace aproximadamente 45000 años, y a partir de aquí se habrían dispersado hacia Europa y Asia. De esta forma, todos los haplotipos mitocondriales actuales tendrían que hacer coalescencia en los ADNmt fundadores que llegaron a estas regiones.
- **Ruta Sur:** esta hipótesis propone una dispersión por la costa sur de Asia, procedente del este de África, hace unos 60000 años aproximadamente. A partir de aquí habrían surgido el resto de los ADNmt que podemos encontrar hoy en día (Macaulay et al. 2005).

- **Dos oleadas de dispersión:** según esta hipótesis, la colonización del hombre moderno se explicaría mediante las dos rutas anteriores, de forma que primero (hace unos 60000 años) hubo una ola de dispersión por la costa sur de Asia, dando lugar a los tipos de ADNmt del sur y este de Asia, y de Oceanía, y posteriormente (hace 45000 años) otra corriente de dispersión tuvo lugar por el norte de África hasta Oriente Medio, desde donde surgieron movimientos migratorios hacia el norte de Asia y hacia Europa (Maca-Meyer et al. 2001).

El ADNmt también ha contribuido al conocimiento de la colonización del Nuevo Mundo (Perego et al. 2009; Torroni et al. 1992), las migraciones iniciales a Nueva Guinea y Australia (Ingman and Gyllensten 2003) y el asentamiento en Europa (Torroni et al. 2006).



**Figura 6:** Representación gráfica de las rutas generales de dispersión de los linajes mitocondriales. Imagen modificada de <http://www.familytreedna.com>.

#### 4.4 Limitaciones del ADNmt

A pesar de que este genoma ha sido y continúa siendo de gran utilidad en el área de la genética de poblaciones, no hay que olvidar que, por sus características propias, se trata de un único marcador y además uniparental. Esto significa que las conclusiones extraídas de su análisis deben ser interpretadas en la medida que

este locus lo permite. El hecho de ser un único marcador y de no presentar recombinación, facilita su análisis e interpretación de los datos, pero también hay que tener presente que no representa la historia global de las poblaciones humanas sino una historia parcial del componente femenino de dichas poblaciones.

Por otra parte, el hecho de que la tasa de mutación varíe considerablemente entre posiciones específicas dificulta la obtención de una estima de la tasa de sustitución promedio para toda la molécula, necesaria para el cálculo de la edad de coalescencia entre dos ramas de la filogenia.

## **5 Aplicaciones forenses de la variabilidad del genoma mitocondrial**

### **5.1 Generalidades**

La genética forense consiste en la aplicación de los conocimientos de la genética a la resolución de problemas judiciales, y dentro de ella pueden englobarse diferentes pericias: investigación biológica de parentesco, estudio de vestigios biológicos de interés criminal e identificación genética. Esta área de conocimiento ha avanzado de forma considerable desde sus comienzos, tanto en lo que se refiere al tipo de marcadores analizados como la cantidad/calidad de la muestra requerida.

Inicialmente, la genética forense se basaba en el estudio de marcadores genéticos clásicos, antígenos sanguíneos (sistemas AB0, MNS, Rh,...) y proteínas séricas. En la década de los setenta, el análisis de los grupos enzimáticos, obtenidos a partir de hemólisis de eritrocitos, experimentó un gran desarrollo. Sin embargo, estos sistemas aportaban poca información ya que su variabilidad es limitada, por lo que el descubrimiento del sistema HLA (antígenos de los leucocitos humanos) supuso un gran avance para la práctica forense, ya que era el sistema antigénico más polimórfico conocido hasta esa fecha.

Wyman y White (1980) introdujeron el concepto de identificación genética a través del estudio de RFLPs (*Restriction Fragment of Length Polymorphism*) y el empleo de estos marcadores genéticos sustituyó en gran medida el uso de los marcadores proteicos clásicos. Las ventajas que aportaba el análisis del ADN eran su mayor variabilidad y estabilidad molecular, ya que las proteínas son más sensibles a la degradación. Poco más tarde, Jeffreys et al. (1985) describieron un método de identificación individual que denominaron “*DNA fingerprinting*” o huella genética, con importantes aplicaciones en la investigación biológica de la paternidad y en criminalística. Utilizando sondas multilocus analizaban minisatélites situados en diferentes regiones del genoma, obteniendo un patrón de bandas prácticamente específico para cada individuo. Debido a su difícil estandarización e interpretación, se pasaron a analizar RFPLs mediante sondas unilocus, por lo que se necesitaban varias sondas para conseguir el mismo poder de discriminación.

Poco después del descubrimiento de la huella genética, la introducción de la PCR permitió el análisis de *loci* hipervariables (minisatélites y microsatélites) a partir de vestigios biológicos que no podían ser analizados hasta el momento, debido a la limitada cantidad de ADN que podía obtenerse en la mayoría de los escenarios forenses.

## **5.2 Tipo de marcadores genéticos**

El estudio de los polimorfismos de ADN se ha centrado mayoritariamente en el análisis de marcadores nucleares, especialmente aquellos situados en cromosomas autosómicos; sin embargo para situaciones concretas (agresiones sexuales) el análisis de polimorfismos de cromosomas sexuales es de gran utilidad. Hoy en día, los marcadores autosómicos de elección son los STRs, aunque en muestras donde no se ha podido obtener un perfil satisfactorio, los SNPs se han posicionado como marcadores alternativos de gran importancia (Musgrave-Brown et al. 2007; Sánchez et al. 2006).

En ocasiones, especialmente en contextos forenses, las muestras han estado expuestas a condiciones tan adversas que el análisis del ADN nuclear es inviable debido a su alto grado de degradación o a su escasa cantidad, a veces incluso

inexistente (pelos sin bulbo). En estos casos, el ADNmt representa prácticamente la única herramienta genética disponible ya que, gracias a sus características, aumenta las probabilidades de éxito del análisis.

El genoma mitocondrial puede aportar información de gran utilidad, y en algunas ocasiones suficiente (junto con otras evidencias), para el establecimiento de la identidad o fuente de un determinado espécimen biológico, jugando un papel crucial en la identificación forense. Un ejemplo conocido de la utilidad de este marcador uniparental en el campo forense lo constituye la identificación de los restos del Zar Nicolás II y su familia, los últimos miembros de la dinastía Romanov, asesinados en 1918 a raíz de la revolución Bolchevique (Coble et al. 2009; Gill et al. 1994).

### 5.3 Utilidad y limitaciones del ADNmt

Son varias características de este pequeño genoma las responsables de su gran utilidad a la hora de analizar muestras degradadas, pelos sin bulbo, material esquelético parcialmente degradado o restos fósiles muy antiguos:

- **Alto número de copias:** frente a las dos copias del genoma nuclear y una copia del cromosoma Y.
- Su **localización en el citoplasma** permite el análisis de este genoma en aquellas células que han perdido el núcleo. El principal ejemplo en la práctica forense son los pelos sin bulbo. A lo largo del pelo, su composición va variando ya que se trata de un tejido vivo que se va queratinizando a medida que crece, por ello las células de la zona distal pierden su núcleo y su ADNn.
- La **elevada tasa de mutación** del ADNmt en comparación con el ADNn ha generado un considerable número de polimorfismos en relación al pequeño tamaño del genoma mitocondrial, lo que aumenta el poder de discriminación.
- La **herencia uniparental y la ausencia de recombinación** del ADNmt, permiten la comparación de miembros de un linaje que comparten un ancestro común por vía materna.

- Su **pequeño tamaño** facilita la automatización de su análisis, que suele consistir en la amplificación mediante PCR y la secuenciación automática de pequeñas regiones hipervariables y considerablemente informativas (HVI y HVII). Además, en los últimos años ha surgido un gran número de nuevas tecnologías para el genotipado de SNPs, lo que permite aumentar el poder de discriminación del genoma mitocondrial.

Sin embargo, las mismas características que hacen que este genoma presente ventajas en situaciones concretas, hacen que en aquellos casos en los que se disponga de ADNn, éste sea el genoma de elección:

- Como consecuencia de heredarse por vía materna y no recombinar, el ADNmt sólo aporta información del linaje materno, pero no permite individualizar dentro de éste.
- El poder de discriminación no es comparable al del genoma nuclear.
- Otra limitación importante es la falta de bases de datos que proporcionen estimas fiables en relación a la frecuencia de un perfil en la población, o a la tasa de mutación sitio-específica. A pesar de ello, existen revisiones con listas de las principales posiciones *hotspots* que pueden servir de ayuda en el contexto forense (Salas et al. 2007; Soares et al. 2009).
- La elevada tasa de mutación de este genoma también puede provocar la aparición de mutaciones somáticas que pueden oscurecer la comparación entre la muestra dubitada y la indubitada, por lo que una única diferencia no supone una evidencia concluyente de exclusión (Paneto et al. 2007). De hecho, se han detectado diferencias en la proporción de las variantes alélicas de una posición heteroplásmica en distintas secciones de un mismo pelo (Salas et al. 2001a), llegando alguna a suponer un cambio de base completo (Tully et al. 2004). Por ello, siempre que sea posible, es recomendable analizar más de una muestra de cabello para confirmar el perfil genético y tener en cuenta las tasas de mutación sitio-específicas.

Es necesario destacar la importante labor de grupos como la ISFG o el grupo EDNAP en el ámbito forense, tanto en lo que se refiere al desarrollo de recomendaciones (Carracedo et al. 2000; Parson and Bandelt 2007; Tully et al.



2001) como la organización de ejercicios colaborativos inter-laboratorio (GHEP-ISFG).

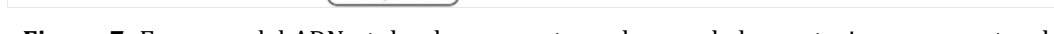
## 6 Aplicaciones clínicas de la variabilidad del genoma mitocondrial

### 6.1 Enfermedades mitocondriales

Se puede considerar el trabajo de Luft et al. (1962) como el comienzo de lo que hoy se entiende por “enfermedades mitocondriales”, ya que por primera vez se asocia un estado clínico complejo con un defecto en el control respiratorio mitocondrial. Actualmente ya se conoce un número bastante elevado de desórdenes mitocondriales, clínicamente heterogéneos, que surgen como resultado de la disfunción de la cadena respiratoria mitocondrial.

La mitocondria tiene entre 1000 y 3000 proteínas, implicadas en rutas metabólicas importantes, además de la cadena de transporte de electrones y la fosforilación oxidativa. Estas proteínas están codificadas en su mayoría por el genoma nuclear, sin embargo, también la pequeña molécula de ADN mitocondrial codifica para 13 proteínas mitocondriales, por lo que mutaciones en cualquiera de los dos genomas pueden provocar desórdenes mitocondriales. Esta circunstancia condiciona que las enfermedades mitocondriales sigan un patrón de transmisión particular, de forma autosómica, para las alteraciones que tienen lugar en el ADNn, y herencia materna para mutaciones del ADNmt (Herrnstadt and Howell 2004). Estas enfermedades mitocondriales suelen afectar a varios órganos, principalmente al sistema nervioso y muscular. Los desórdenes mitocondriales pueden aparecer a cualquier edad, aunque en general se presentan en edades tempranas cuando la mutación causante es nuclear y durante la adolescencia o edad adulta cuando se trata de una mutación mitocondrial (Patrick Chinnery; [http://www.ncbi.nlm.nih.gov/bookshelf/br.fcgi?book=gene&log\\$disease\\_name&part=mt-overview](http://www.ncbi.nlm.nih.gov/bookshelf/br.fcgi?book=gene&log$disease_name&part=mt-overview)).

En 1988 se asociaron por primera vez **mutaciones en el ADNmt** con enfermedades humanas (Holt et al., 1988) y desde entonces ha ido creciendo el



- MERRF: epilepsia mioclónica con fibras rojas rasgadas.
- MELAS: encefalomiopatía mitocondrial con acidosis láctica y accidentes vasculares cerebrales.
- Síndrome Kearns-Sayre (KSS): oftalmoplejía externa progresiva, retinitis pigmentaria.
- NARP: neuropatía sensitivo-motora, ataxia y retinitis pigmentaria.
- CPEO: oftalmoplejía externa progresiva crónica.
- LHON: atrofia óptica hereditaria de Leber.
- MDS: síndrome de depleción de ADNmt.

La clasificación de las enfermedades mitocondriales se puede realizar desde el punto de vista bioquímico o bien según el defecto genético causante de la patología.

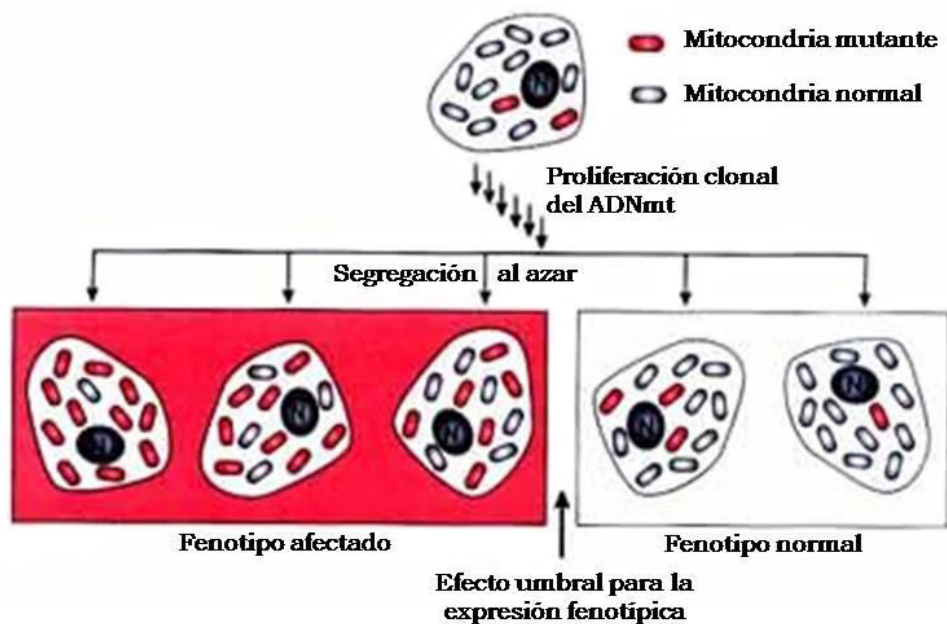
CLASIFICACIÓN DE ENFERMEDADES MITOCONDRIALES	
GENÉTICA	POR DEFECTOS DEL METABOLISMO ENERGÉTICO
<b>Alteraciones del ADNmt</b> Alteración de la síntesis proteica mitocondrial Alteración de genes que codifican proteínas de la cadena respiratoria <b>Alteraciones del ADNn</b> Alteraciones de los genes que codifican proteínas mitocondriales Alteraciones en la importación de proteínas mitocondriales Alteraciones en la comunicación intergenómica Alteraciones en el medio lipídico Alteraciones en la motilidad/fusión/fisión mitocondrial	Defectos de la oxidación de los ácidos grasos Defectos del metabolismo del piruvato Defectos del ciclo de Krebs Defectos en el acoplamiento oxidación-fosforilación Defectos de la cadena respiratoria mitocondrial

**Tabla 4:** Clasificación de las enfermedades mitocondriales en función del defecto genético o bioquímico causante de la patología.

En lo relativo a desórdenes causados por alteraciones en el ADNmt, a pesar de que existen una serie de síndromes bien definidos, su característica principal es la heterogeneidad en sus manifestaciones. Esta variabilidad viene en parte

condicionada por los siguientes fenómenos, íntimamente relacionados con las características de las mitocondrias, heteroplasmia, segregación mitótica y efecto umbral. Las mutaciones patogénicas son normalmente **heteroplásmicas**, con poblaciones de moléculas “salvajes” y “mutantes” presentes en la misma célula. Durante la mitosis, ambos tipos de ADNmt segregan al azar en las células hijas (**segregación mitótica**), con lo que la proporción de moléculas “mutantes” que pasarán a la siguiente generación de células puede variar mucho, afectando a la expresión de la enfermedad.

La expresión fenotípica de una mutación patogénica mitocondrial no sigue las reglas de herencia mendeliana y depende en gran medida de las proporciones de ADNmt normal y mutado existentes en un tejido. El porcentaje de mitocondrias afectadas necesarias para que se desarrolle la patología varía en función de cada tejido y sus necesidades energéticas. Se entiende por **efecto umbral** a la proporción mínima de ADNmt mutado necesaria para alterar el metabolismo oxidativo a un nivel suficiente para producir la disfunción de un determinado órgano o tejido.



**Figura 8:** Representación gráfica de la segregación mitótica y del efecto umbral para la expresión fenotípica de una mutación del ADNmt.

Por otra parte, más de 150 mutaciones puntuales e innumerables reordenamientos a gran escala han sido asociados con una gran variedad de

enfermedades. La justificación se halla en la ubicuidad de las mitocondrias en todos los tejidos humanos y en su principal función, la obtención de la mayor parte de la energía celular (Dimauro and Davidzon 2005). A pesar de ello, hay que tener presente que el ADNmt se caracteriza por presentar una tasa de mutación considerablemente elevada, y la gran mayoría de los cambios son polimorfismos frecuentes en las poblaciones humanas, muchos de los cuales caracterizan los linajes mitocondriales. Por ello, se vio la necesidad de probar la patogenicidad de las mutaciones nóveles encontradas para lo cual se establecieron una serie de criterios que dichas alteraciones deben cumplir (Dimauro and Davidzon 2005):

1. La mutación no debe estar presente en individuos sanos del mismo grupo étnico.
2. Debe alterar una posición conservada evolutivamente y por tanto funcionalmente importante.
3. Debe causar una o varias deficiencias enzimáticas en la cadena respiratoria en los tejidos afectados o defectos en la síntesis proteica mitocondrial.
4. Debe haber una correlación entre el grado de heteroplasmía y la severidad clínica así como con la patología celular.

Otros autores propusieron un sistema de puntuación basado en evidencias funcionales (defectos bioquímicos, estudios funcionales en líneas celulares), grado de conservación evolutiva, segregación en familias, detección en estudios independientes y grado de heteroplasmía para mutaciones detectadas en genes mitocondriales de ARN transferentes (McFarland et al. 2004) y en los siete genes que codifican para subunidades del complejo I (Mitchell et al. 2006).

## **6.2 Estudios de asociación caso-control**

Recientemente se ha sugerido que los distintos haplotipos del ADNmt pueden modular la cadena respiratoria, determinando así en cierta medida la fisiología general del individuo y de esta forma aportándole una predisposición o protección frente a determinadas enfermedades.

Los estudios de asociación basados en haplogrupos mitocondriales se emplean para definir el papel que juega la variabilidad del genoma mitocondrial en

relación al riesgo de padecer enfermedades complejas. En concreto, las hipótesis propuestas se basan en la principal función mitocondrial, la producción de energía, y en su papel en la vía intrínseca de la apoptosis.

Las variantes o polimorfismos del ADNmt podrían modificar sutilmente la eficiencia de la cadena respiratoria desacoplando el equilibrio entre producción de especies reactivas de oxígeno (ROS) durante el transporte de electrones y la producción de energía. Aquellas variantes que aumentasen la producción de ATP liberarían una menor proporción de energía en forma de calor e incrementarían la formación de ROS, y viceversa. De hecho, algunos autores propusieron que determinados haplogrupos provocan una mayor producción de calor como consecuencia de una fosforilación oxidativa ineficiente, y esto habría determinado la selección adaptativa de los diferentes linajes mitocondriales a factores ambientales como el clima (Mishmar et al. 2003). Sin embargo, estudios basados en medidas directas de la eficiencia del acoplamiento de la cadena respiratoria no encontraron diferencias significativas entre distintos haplogrupos mitocondriales (Elson et al. 2007).

En estos estudios de asociación, casos y controles son clasificados en los principales haplogrupos mitocondriales y se compara la distribución de los linajes en ambos grupos. Una diferencia significativa en las frecuencias de los haplogrupos, o de los SNPs diagnóstico, es interpretada como un indicio de que el “ambiente” mitocondrial posee un efecto en la expresión o penetrancia del desorden clínico. Una posibilidad sería que una o varias posiciones que definen el haplogrupo pudiesen explicar, al menos parcialmente, la susceptibilidad a la enfermedad. Esto se conoce como hipótesis enfermedad común-variante común o CD-CV (*common disease-common variant*), según la cual muchas de las enfermedades complejas no son causadas por un único factor de riesgo de gran efecto (ya sea genético o ambiental), sino por la acumulación de numerosos factores de pequeño efecto en el riesgo a padecer la enfermedad. Otra posibilidad sería la existencia de alguna mutación asociada a un haplogrupo en la población de estudio. El efecto de variantes individuales es difícil de evaluar porque muchas de ellas están asociadas a más de un linaje.

En los últimos años se ha debatido mucho sobre el papel de la disfunción mitocondrial en relación con un gran número de desórdenes de etiología compleja. El espectro de enfermedades tratadas en este tipo de estudios de asociación es muy amplio, desde Parkinson, Alzheimer, desorden bipolar, esquizofrenia, cáncer, sepsis..., hasta movilidad espermática y envejecimiento. El número de trabajos en el que se asocian SNPs o haplogrupos mitocondriales con enfermedades complejas es cada vez mayor, pero también han surgido revisiones que cuestionan estas asociaciones (Herrnstadt and Howell 2004; Kong et al. 2006; Raule et al. 2007; Salas et al. 2005b). La problemática general es la ausencia de replicación de dichas asociaciones en estudios independientes y la ausencia de ensayos en los que se demuestren las asociaciones defendidas, lo que pone en duda la veracidad de este tipo de asociaciones. Esto es especialmente evidente para enfermedades como el Alzheimer (Chinnery et al. 2000; van der Walt et al. 2004) o el Parkinson (Pyle et al. 2005; Ross et al. 2003).

### **6.2.1 Estudios de asociación entre variantes del ADNmt y cáncer de mama**

El cáncer de mama es el resultado de la multiplicación excesiva, crecimiento desordenado y falta de diferenciación de células de las glándulas mamarias, dando lugar al desarrollo de tejido canceroso. Es el cáncer más diagnosticado en mujeres y el segundo tipo de cáncer de mayor mortalidad entre este sector. Tanto factores ambientales (terapia de sustitución hormonal, edad de menopausia, obesidad,...) como genéticos influyen en el riesgo de una mujer a desarrollar cáncer de mama. Actualmente se conocen algunos genes de alto riesgo para el cáncer de mama (*BRCA1*, *BRCA2*, *PTEN*, *TP53*,...) y otros que poseen un riesgo moderado (*CHEK2*, *TGF $\beta$ 1*, *CASP8*,...), sin embargo estos genes sólo explican un 25% de los cánceres familiares y entre 5-10% de los casos de cáncer de mama de la población general. Debido al pequeño porcentaje de casos explicados por mutaciones en los genes mencionados, se piensa que esta enfermedad puede ajustarse a un modelo poligénico, donde varios genes serían los causantes de la enfermedad. Por ello continúa la búsqueda de estos factores genéticos, y una de las estrategias más populares son los estudios de asociación.

En los últimos años, se ha publicado un número considerable de estudios de asociación que intentan relacionar la variabilidad del genoma mitocondrial con el riesgo de padecer cáncer. Las hipótesis propuestas para explicar esta posible asociación, al igual que para otras enfermedades complejas, tienen su base en el metabolismo energético y la muerte celular programada. Esta idea nace en 1926 cuando Warburg establece que las células cancerosas presentan deficiencias respiratorias y sugiere que pueden ser consecuencia de daños en la mitocondria y una función afectada de la misma. Este fenómeno, conocido como efecto Warburg y caracterizado por el aumento de la glicólisis para producir ATP y por un desequilibrio redox, ha sido el punto de partida para dirigir la búsqueda de los defectos genéticos causantes del cáncer hacia el genoma mitocondrial (Carew and Huang 2002).

Actualmente la hipótesis propuesta de forma generalizada consiste en que las variantes del ADNmt modificarían sutilmente la eficiencia de la cadena respiratoria de tal forma que se desacoplaría el equilibrio entre producción de especies reactivas de oxígeno (ROS) y la capacidad de detoxificación de las enzimas mitocondriales. Las especies reactivas de oxígeno provocan un mayor estrés oxidativo para la célula, estado que a su vez está relacionado con el daño a macromoléculas celulares (oxidación de proteínas y peroxidación de lípidos) y a sus sistemas genéticos (ADNn y ADNmt), favoreciendo así la consiguiente disfunción mitocondrial. En último término la consecuencia de la variabilidad mitocondrial sería el desarrollo del cáncer (Carew and Huang 2002). En el caso del cáncer de mama se ha propuesto que pueda tener lugar una mayor producción de ROS a nivel local como consecuencia del metabolismo de hormonas (estradiol), lo que podría constituir un efecto adicional (Canter et al. 2005).

En 2005 se publicó un trabajo (Canter et al. 2005) en el que estudiaron una variante del ADNmt (10398A) en relación con el riesgo de padecer cáncer de mama en mujeres afroamericanas. Analizaron 654 casos de cáncer de mama invasivo y 605 controles y encontraron asociación estadísticamente significativa ( $p$ -valor=0'013). No pudieron confirmar la asociación en una muestra de mujeres "blancas", por lo que los autores propusieron que el alelo A de este polimorfismo



solamente confería riesgo a padecer la enfermedad a mujeres afroamericanas mediante interacción con otros factores genéticos o ambientales.

En un trabajo posterior (Bai et al. 2007), se analizaron 156 casos de cáncer de mama familiar y 260 controles, ambos grupos formados por mujeres americanas de ascendencia europea, para los cuales se genotiparon determinadas variantes del genoma mitocondrial, algunas de ellas diagnóstico de los principales haplogrupos mitocondriales europeos. Encontraron que tres mtSNPs estaban asociados significativamente con la enfermedad, dos de ellos (9055A y 16519C) confiriendo susceptibilidad frente a la misma, mientras que el 3197C poseía un efecto protector. Mediante un modelo de regresión logística, detectaron que el 13708A aportaba protección frente al cáncer de mama familiar, mientras que el 10398G confería un mayor riesgo. Al evaluar los haplogrupos asociados, encontraron que los individuos pertenecientes al haplogrupo K presentaban mayor riesgo, mientras que los individuos pertenecientes al haplogrupo U presentaban protección.

Recientemente ha sido publicado un estudio (Covarrubias et al. 2008) cuyo objetivo era demostrar que la interacción entre diferentes variantes del genoma mitocondrial modifica el riesgo de padecer cáncer de mama. Para ello, emplearon los datos del trabajo anterior (Bai et al. 2007), si bien esto no se dice de forma explícita, y evaluaron las interacciones dos a dos mediante un modelo de regresión logística. Encontraron que el 10398G, de forma individual, representa un aumento de riesgo de padecer cáncer de mama, al igual que la interacción 12308G-10398G (estadísticamente significativa), mientras que 12308G-10398A presentan un efecto protector. Además, parece que el haplogrupo T (12308A-10398A, entre otros polimorfismos) no tiene efecto sobre el riesgo a padecer cáncer de mama, mientras que el haplogrupo U (12308G-10398A) parece poseer un efecto protector. Estos estudios requieren, sin embargo, una reflexión sobre el verdadero significado de una interacción en un contexto haplotípico y, dado que se tratan de los mismos datos, no deben considerarse como evidencias independientes a favor de la asociación del ADNmt con cáncer de mama (comunicación personal de Antonio Salas; datos todavía sin publicar).

En los últimos años se han publicado nuevos trabajos intentando descifrar el papel que juega el mtSNP A10398G en relación al cáncer de mama. Setiawan et al.

(2008), en un estudio de tamaño muestral considerable (1457 casos y 978 controles, distribuidos en tres estudios), no detectaron asociación entre el alelo A y la enfermedad en tres poblaciones de afroamericanas. Czarnecka et al. (2009) genotiparon este SNP en población polaca (44 casos y 100 controles) mediante PCR-RFLP y encontraron una alta asociación estadística entre el alelo G y el riesgo a padecer esta enfermedad. Pezzotti et al. (2009) no detectaron asociación entre el SNP y el riesgo a padecer cáncer de mama en ninguno de los dos estudios (NHS: 1561 casos y 2209 controles; WHS: 678 casos y 669 controles). Sin embargo, analizando la posible interacción con el consumo de alcohol, encontraron que las mujeres no bebedoras presentan menor riesgo de padecer cáncer de mama (21%) si presentan el alelo G en comparación con el alelo A, y que el riesgo aumenta un 56% si portadoras del alelo G consumen alcohol.

En resumen, no se han podido replicar adecuadamente las asociaciones detectadas entre el ADNmt y el cáncer de mama, por lo que es difícil sacar conclusiones sobre el papel que juega el genoma mitocondrial en el desarrollo o progresión del cáncer de mama.

### **6.2.2 Estudios de asociación entre variantes del ADNmt y esquizofrenia**

La esquizofrenia es un desorden psiquiátrico con una prevalencia del 0'72%, que suele aparecer en la adolescencia tardía y que habitualmente cursa como una enfermedad crónica. Clínicamente es muy heterogénea entre individuos, con una gran variabilidad en la presentación sintomática, e incluso la sintomatología de un individuo puede variar a lo largo de su vida. La heredabilidad de esta enfermedad es del 80% aproximadamente, de forma que los factores de riesgo genéticos juegan un papel importante en el desarrollo de esta enfermedad. La mayoría de los estudios genéticos relacionados con esta enfermedad se basan en análisis de asociación y de ligamiento centrándose en regiones cromosómicas. A pesar de que algunos de ellos han dado resultados positivos y bien replicados, como es el caso de la implicación del sistema dopaminérgico y la posible interacción con otros neurotransmisores, aún no se ha encontrado la causante genética de la esquizofrenia (Costas et al. 2009).

Como consecuencia de que los *loci* del genoma nuclear identificados hasta el momento solamente explican una pequeña proporción del riesgo genético de la esquizofrenia, algunos autores han empezado a mirar hacia el otro genoma celular como posible responsable, al menos en parte, de este desorden psiquiátrico. La hipótesis de la implicación del ADNmt en la susceptibilidad genética para la esquizofrenia está apoyada desde varias perspectivas, algunas de las cuales también se proponen para otros desórdenes psiquiátricos (Ben-Shachar 2002; Martorell et al. 2006):

- Algunas evidencias sugieren una mayor transmisión materna de este desorden psiquiátrico, ya que el riesgo de tener descendencia con esquizofrenia es mayor si la madre padece la enfermedad que si la padece el padre. Además, el hecho de que la prevalencia de la esquizofrenia sea del 1%, a pesar de que la capacidad reproductiva de los pacientes está disminuida (especialmente en varones) y que la mortalidad sea más elevada en jóvenes, ha llevado a sugerir que la transferencia de genes de susceptibilidad pueda involucrar a la línea materna (Doi and Hoshi 2007).
- Importancia de la mitocondria en la función neuronal. Las neuronas tienen una membrana celular excitable, polarizada, que permite la entrada de iones desde el espacio extracelular, pudiendo responder así a estímulos exteriores. El mantenimiento de este gradiente es energéticamente muy costoso, por lo que se ha sugerido que un estado energético anormal, como consecuencia de la disfunción mitocondrial, podría causar alteraciones en la función de las neuronas, en su plasticidad y en el circuito cerebral. Esto podría explicar las aberraciones cognitivas y de comportamiento características de la esquizofrenia.
- Hipoplasia mitocondrial: deformación y reducción del número de mitocondrias
- La esquizofrenia parece compartir varias características con las enfermedades mitocondriales:
  - Patogénesis en cerebro: metabolismo energético disminuido en el cerebro.
  - Expresión alterada de genes relacionados con la mitocondria.
  - Severidad variable.

- Detección de síntomas psiquiátricos en pacientes con enfermedades mitocondriales.
- Interacción entre la dopamina, uno de los principales factores etiológicos de la esquizofrenia, y la respiración mitocondrial. Parece que el tratamiento con antipsicóticos, que bloquean los receptores de dopamina D2, reduce la actividad del complejo I de la cadena de transporte de electrones.

Marchbanks et al. (2003) genotiparon una variante del gen *ND4* (12027T>C) del ADNmt, previamente detectada en heteroplasmía en un paciente con esquizofrenia, en 181 casos de esta enfermedad y 184 controles mediante PCR-RFLP. Encontraron una elevada frecuencia de esta variante en heteroplasmía en pacientes (81/181), a pesar de que la frecuencia también era considerablemente alta en el grupo control (43/184). La frecuencia de esta posición heteroplásmica también variaba en función del sexo del paciente, y sugirieron que la población femenina es menos sensible a la presencia de una variante en términos de predisposición para la esquizofrenia, ya que el 38% de los controles femeninos presentaban la heteroplasmía.

Poco después, Bandelt et al. (2005c), a raíz de las inconsistencias detectadas en el trabajo anterior (Marchbanks et al. 2003), hicieron hincapié en la importancia de tener en cuenta una serie de medidas para interpretar adecuadamente los datos en un estudio de caso-control antes de concluir la existencia de asociación entre una mutación y una enfermedad. En este tipo de literatura los principales problemas son artefactos de genotipado (digestión enzimática parcial, lectura de secuencias en una sola dirección), ausencia de evaluación filogenética de las mutaciones, no tener en cuenta criterios de patogenicidad o ausencia de estudios funcionales.

En un trabajo del 2006 (Martorell et al. 2006) se secuenció el genoma mitocondrial completo de seis pacientes con aparente transmisión materna de esquizofrenia, para ver si eran portadores de mutaciones que pudieran contribuir a la disfunción mitocondrial descrita en este desorden. Encontraron tres mutaciones sin sentido (*missense*), que se analizaron mediante RFLPs en 95 controles y observaron que solamente la variante T12096A estaba conservada evolutivamente y podía suponer un cambio en la estructura de la subunidad ND4

del complejo I. Concluyen que variantes en subunidades del complejo I pueden estar asociadas con desórdenes psicóticos o que la presencia o acumulación de determinadas variantes puede contribuir a la base genética de síntomas específicos de la esquizofrenia.

De nuevo, Bandelt et al. (2007), a partir del trabajo de Martorell et al. (2006) proponen evaluar detenidamente las posibles causas de una asociación positiva falsa antes de interpretar prematuramente los datos. Es fundamental eliminar la posibilidad de contaminación cruzada secuenciando fragmentos ampliamente solapantes en ambas direcciones, tener precaución con los resultados de RFLPs y revisar las variantes halladas desde una perspectiva filogenética, siempre comparándolas con rCRS.

En 2007 se publica un trabajo (Magri et al. 2007) en el que se pretendía determinar si existían diferencias en la distribución de haplogrupos mitocondriales entre 190 controles y 142 pacientes con esquizofrenia. No encontraron asociación estadística entre ningún haplogrupo y el riesgo a padecer este desorden, aunque sí observaron que individuos pertenecientes al haplogrupo J-T presentaban una edad de aparición de la enfermedad más temprana. En este trabajo no detectaron la variante T12096A hallada en el trabajo de Martorell (2006) en ninguno de los pacientes que parecían presentar herencia materna de la enfermedad (N=13).

En un estudio similar (Amar et al. 2007) con población árabe de Israel, observaron que el haplogrupo HV, el linaje más prevalente en esta población, se hallaba sobre-representado en el grupo de pacientes con esquizofrenia, aportándoles un riesgo relativo de 1'8.

Bamne *et al.* (2008), en una primera etapa, secuenciaron el genoma completo de 180 casos de esquizofrenia y 180 controles en *pools* de ADN, y encontraron dos polimorfismos (G13368A y G13708A) del gen *ND5* que podrían sugerir una tendencia en las frecuencias alélicas. En una segunda etapa, secuenciaron el gen *ND5* en el primer conjunto de muestras, y en un segundo *set* compuesto por 244 casos y 508 controles. No encontraron diferencias en las frecuencias alélicas de ninguno de los dos polimorfismos iniciales ni de otras variantes no sinónimas.

Los autores de un estudio reciente (Ueno et al. 2009) analizaron el genoma mitocondrial completo de 93 casos de esquizofrenia y examinaron la distribución de los principales linajes mitocondriales entre estos pacientes y 784 controles para evaluar la posible asociación entre variantes raras o comunes del ADNmt y la patogénesis de la esquizofrenia. Encontraron tres variantes homoplásmicas raras de la subunidad 6 de la ATP sintasa conservadas evolutivamente (T8843C, G8902A y T8945C) que podrían ser ligeramente deletéreas y tres variantes heteroplásmicas nuevas (G1227A, T5578C, G13418A) potencialmente patogénicas. Sin embargo, no encontraron una distribución de haplogrupos significativamente diferente entre casos y controles.

Los trabajos que defienden asociaciones entre el ADNmt y diferentes enfermedades complejas superan mucho en número a aquellos estudios que muestran mayor escepticismo, sin embargo, es necesario ser cautos a la hora de formular conclusiones. Algunos autores pretenden llamar la atención sobre la enorme importancia de interpretar adecuadamente los resultados obtenidos en un estudio de asociación (Bandelt et al. 2007; Bandelt et al. 2009; Bandelt et al. 2005c; Salas et al. 2005b).

### **6.2.3 Enfermedades infecciosas**

La variabilidad del genoma mitocondrial se ha asociado recientemente con varias enfermedades infecciosas, entre ellas sepsis severa, SIDA o tuberculosis.

La sepsis o síndrome de respuesta inflamatoria sistémica (SRIS) es un síndrome clínico caracterizado por el fallo multiorgánico tras una infección microbiológica. Es un desorden muy común en el ámbito hospitalario y es una de las principales causa de muerte en pacientes de Unidades de Cuidados Intensivos (UCI). Existen numerosas publicaciones que sugieren la relación entre la disfunción mitocondrial y la sepsis, basándose en:

- Parece que la disfunción multiorgánica en la sepsis severa está relacionada con una hipoxia citopática (disoxia), es decir, un deterioro en la respiración celular a pesar de los niveles adecuados de oxígeno a nivel tisular.

- Se piensa que para una respuesta inmune efectiva ante una sepsis se requiere una gran cantidad de energía.
- Se ha propuesto que una posible causa de sepsis y shock séptico pueda ser una respuesta inmune e inflamatoria excesiva caracterizada por altos niveles de ROS, óxido nítrico (NO) y citoquinas. Las consecuencias de esto serían daño al endotelio vascular, respiración tisular disminuida y disfunción de la respiración mitocondrial (menos ATP y consumo de O<sub>2</sub>).
- Se detectaron anomalías estructurales y bioquímicas en mitocondrias de pacientes sépticos, y reducida actividad de las enzimas mitocondriales en tejido muscular.

Posteriormente se asoció la variabilidad del genoma mitocondrial y la sepsis, basándose en supuestos cambios en las actividades de los complejos de la cadena respiratoria que presentarían los distintos linajes mitocondriales.

Baudouin et al. (2005) llevaron a cabo un estudio para determinar si el haplogrupo H influía en la supervivencia tras sufrir sepsis. Para ello analizaron la distribución de los haplogrupos mitocondriales en 150 pacientes con infecciones severas y en 542 controles. Encontraron que el haplogrupo H, el más frecuente en poblaciones europeas, es un predictor independiente y robusto del pronóstico durante sepsis severa, ya que pacientes con este haplogrupo presentaban una probabilidad superior de sobrevivir. También presentaban una temperatura basal estadísticamente superior a individuos pertenecientes a otros haplogrupos, lo que concluyeron que podría ser una ventaja adaptativa y una posible explicación para la gran expansión de este haplogrupo a pesar de su relativa juventud. Sin embargo, esta hipótesis sobre las diferencias bioenergéticas entre los haplogrupos T y H no fue confirmada empíricamente (Amo et al. 2008).

Yang et al. (2008), en un estudio sobre la distribución de haplogrupos mitocondriales en 181 pacientes Han admitidos en la UCI y 570 controles, concluyen que el haplogrupo R es un fuerte predictor del pronóstico en sepsis severa. No encuentra diferencias significativas en la distribución de los principales subhaplogrupos del haplogrupo R de la población Han. En otro trabajo (Gómez et al. 2009), en el que se genotipó la variante T4216C en 136 pacientes con trauma

severo admitidos en la UCI, encontraron que ser portador del alelo C estaba significativamente asociado con un mayor riesgo de shock séptico o disfunción multiorgánica tras un trauma severo, con respecto al alelo T.

Varios trabajos han asociado también la disfunción mitocondrial con otras enfermedades infecciosas como el síndrome de inmunodeficiencia adquirida (SIDA) o la tuberculosis. Hendrickson et al. (2008) encontraron que los haplogrupos J y U5 son más frecuentes en pacientes infectados por el VIH que desarrollan una rápida progresión de la enfermedad, mientras que los haplogrupos H, UK e IWX parece poseer un fuerte efecto protector frente la progresión del SIDA. Concluyen que la variabilidad del ADNmt entre haplogrupos puede alterar la producción de ATP, ROS o la apoptosis y por eso está correlacionada con la progresión de esta enfermedad. No se ha detectado asociación estadísticamente significativa entre los haplogrupos mitocondriales y la infección de tuberculosis (Houshmand et al. 2007).

La **enfermedad meningocócica** engloba un grupo de síndromes clínicos, de gravedad variable, ocasionados por la bacteria *Neisseria meningitidis*. Entre ellos se pueden distinguir:

- Meningococemia: infección del torrente sanguíneo, que puede variar desde leve a sepsis meningocócica fulminante (infección generalizada). Los síntomas pueden aparecer de forma repentina comenzando por fiebre, decaimiento y aparición de petequias (pequeñas manchas violáceas) en la piel y evolucionar muy rápidamente (fallo circulatorio, fallo multiorgánico, coma), pudiendo ser mortal en pocas horas.
- Meningitis meningocócica: infección de las membranas que circundan el cerebro y la médula espinal que cursa con dolores de cabeza, cuello y espalda, fiebre, náuseas, vómitos y rigidez de nuca.
- Otras manifestaciones: neumonía meningocócica, artritis,...

Atendiendo a las diferencias antigénicas de sus polisacáridos capsulares, se conocen 13 grupos serológicos, aunque la mayoría de las cepas responsables de la enfermedad meningocócica pertenecen a los serogrupos A, B, C, W135 e Y.



Se estima que el 10% de la población sana son portadores asintomáticos de la bacteria, y su reservorio natural es la nasofaringe. La enfermedad ocurre cuando el microorganismo invade la mucosa nasofaríngea y pasan a la circulación sanguínea. La enfermedad meningocócica puede presentarse a cualquier edad, pero es más frecuente en niños entre uno y cinco años, y es una de las principales causas de muerte por infección en los países industrializados. Se cree que la variabilidad en la respuesta a esta infección depende de factores bacterianos, ambientales y del hospedador, pero los mecanismos moleculares y celulares de la sepsis meningocócica se desconocen.

#### 6.2.4 Autismo y desórdenes del espectro autista

El **autismo** es un trastorno que afecta a la comunicación, interacción, imaginación, planificación y reciprocidad emocional. Los síntomas, en general, son incapacidad de interacción social, aislamiento, estereotipias (movimientos incontrolados de alguna extremidad, generalmente las manos), conductas muy rígidas, negación a cualquier cambio o novedad y repetir una acción determinada muchas veces. Puede pasar desapercibido los primeros meses de vida, hasta que se hace evidente la falta de reciprocidad con el interlocutor.

El término “**trastornos del espectro autista**” (ASD, *Autism Spectrum disorders*), hace referencia a un fenotipo más amplio que englobaría tanto el autismo como otros trastornos más leves caracterizados por presentar ciertos rasgos autistas (síndrome de Asperger o el síndrome de Rett). La prevalencia de autismo es de uno o dos casos cada 1000, y para los trastornos del espectro autista es de seis cada 1000. Los niños son más susceptibles a padecer este trastorno que las niñas (4:1).

El espectro autista es un grupo de desórdenes de etiología y presentación clínica heterogénea que afecta al desarrollo del cerebro. Se desconoce la causa que provoca el autismo, pero estudios en familias y con gemelos monocigóticos sugieren que el componente genético juega un papel importante. Se cree que este trastorno tiene una base genética compleja, aunque no está claro si las causantes son mutaciones raras o combinaciones poco frecuentes de variantes genéticas comunes (hipótesis CD-CV). También se ha postulado que determinados factores

ambientales como complicaciones prenatales o perinatales, intoxicación con metales pesados o la vacuna anti-tosferínica podrían afectar el desarrollo del autismo en personas con predisposición genética.

Recientemente en la literatura se ha sugerido la disfunción mitocondrial como posible causa de trastornos del espectro autista. Las principales razones que justifican esta asociación son:

- Aparente herencia materna: parece que las madres de niños con este tipo de trastornos presentan ciertas características autistas.
- Metabolismo energético alterado en estos pacientes e hiperproliferación mitocondrial.
- Mayor frecuencia de enfermedad mitocondrial en pacientes de ASD que en la población general.
- Similitudes con la enfermedad mitocondrial: deficiencias en la cadena respiratoria, fatiga o intolerancia al ejercicio, afectación neurológica (visión, audición, ptosis).

De esta forma, la hipótesis propuesta se basaría en que alteraciones en el ADNmt o en la comunicación intergenómica núcleo-mitocondria alterarían el metabolismo energético modificando así el normal desarrollo neurológico en fases tempranas pudiendo causar trastornos del espectro autista.

Graf et al. (2000) llevaron a cabo un estudio basado en una familia con diferentes desórdenes neurológicos, entre ellos un individuo con autismo y su hermana con síndrome de Leigh. Para estos individuos, se genotipó la posición G8363A además de una serie de mutaciones descritas previamente como patogénicas (A3243G, A3251G, A3260G, A8344G, T8356C, T8851C, T8993C/G, T9176C/G, G14459A, T14709C, A15923G, C15990T). Encuentran la variante G8363A en heteroplasmia (en diferentes proporciones) en varios miembros de la familia, pero ninguna de las otras mutaciones, y concluyen que la presencia de esta posición en heteroplasmia (60%) es el defecto genético responsable del fenotipo autista de este paciente.

Pons et al. (2004) genotiparon la posición A3243G del ADNmt en diferentes tejidos de cuatro pacientes diagnosticados con algún desorden del espectro autista y con historia familiar de enfermedades mitocondriales. El objetivo de este trabajo era determinar la posible asociación entre alteraciones en el ADNmt, concretamente la variante A3243G, y ASD. Encontraron la variante en heteroplasmía en dos pacientes y en otras dos madres (aunque en sus hijos no se detectó el alelo G). Concluyeron que determinadas características autistas pueden ser consecuencia de la expresión de la disfunción mitocondrial en el cerebro, especialmente en las áreas relacionadas con la patología del autismo.

Oliveira et al. (2005) analizaron una cohorte de 102 autistas (sin otro desorden clínico asociado). Analizaron los niveles de lactato en plasma y detectaron que estaban elevados en 14 de ellos. Para once de estos pacientes se determinaron las actividades de los complejos de la cadena respiratoria, y en seis se encontraron deficiencias. En estos seis pacientes se analizaron las posiciones A3243G, T3271C, T3256C, T8356C, G8363A del ADNmt, pero no se detectó ninguna anomalía. Cinco de los cuales se diagnosticaron con desorden mitocondrial definitivo. Los autores concluyeron que las enfermedades mitocondriales son de los desórdenes asociados con más frecuencia al autismo, y probablemente se deban a defectos en genes nucleares, ya que no detectaron mutaciones en el ADNmt.

Weissman et al. (2008) llevaron a cabo una revisión clínica y bioquímica de 25 pacientes que inicialmente fueron diagnosticados de ASD y que posteriormente se determinó que presentaban alguna disfunción en la cadena de transporte de electrones. Encontraron evidencias clínicas poco frecuentes en pacientes idiopáticos de autismo, además de una serie de mutaciones potencialmente patogénicas (T3394C, C10394T, T11809C, T11984C) que aparecían en el 16% de los pacientes, y dos probablemente patogénicas (8% de los pacientes): A3397G (Complejo I) y A4295G (ARNt<sup>Ileu</sup>). Estos autores proponen que un subconjunto de pacientes con ASD presenta alguna disfunción en la producción mitocondrial de energía que los diferencia desde el punto de vista fisiopatológico de los autistas idiopáticos.

En el mismo año, otro trabajo (Kent et al. 2008) evaluó la posible relación entre el ADNmt y el autismo, esta vez analizando la distribución de haplogrupos entre 162 autistas y dos grupos controles. El objetivo de este trabajo era determinar si, mediante cambios sutiles en la cadena respiratoria, los linajes mitocondriales contribuían al riesgo de padecer autismo. A pesar de que encontraron diferencias significativas para un linaje mitocondrial (sin corrección por test múltiple), este hallazgo no se replicaba en el otro grupo control. Por ello estos autores concluyeron que no hay evidencias que indiquen la influencia de los haplogrupos mitocondriales en el desarrollo de este trastorno.

---

## *JUSTIFICACIÓN Y OBJETIVOS*



## Justificación

El estudio de la variabilidad del genoma mitocondrial es de interés en distintos ámbitos de la investigación biomédica, desde la genética de poblaciones, la pericia forense y la clínica. Si bien su estudio ha centrado la atención de una buena parte de la comunidad científica, existen muchos aspectos que aún están por resolver en los distintos campos de la investigación biomédica. Estos aspectos están relacionados con la propia biología del genoma mitocondrial o también con los patrones de diversidad genético poblacional. Cabe mencionar, por ejemplo, cuestiones relacionadas con la estratificación poblacional en lugares geográficos muy diversos. Por ejemplo, existe una amplia mayoría de genetistas que considera que las poblaciones europeas son lo bastante homogéneas para considerar que no existe estratificación suficiente que pueda afectar en los estudios de asociación. Sin embargo, no existen pruebas empíricas suficientes que lo demuestren, sino más bien evidencias indirectas basadas en muestreos poblacionales deficientes.

Quizás sea la genética clínica relacionada con el estudio de enfermedades complejas donde el efecto de la estratificación poblacional está teniendo sus consecuencias más devastadoras: existen varios cientos de trabajos publicados que relacionan determinadas variantes mitocondriales o linajes concretos con patologías comunes (multifactoriales) con conocida base genética. La mayor parte de estos estudios (sino todos) han sido publicados en revistas de impacto medio-alto, pero los estándares de publicación parecen ser más bajos que para otros condicionantes autosómicos, lo que ha hecho que la mayor parte de los estudios presenten una metodología estadística limitada, no estudian réplicas confirmatorias de sus hallazgos de asociación, y no hay estudios de estratificación poblacional. Las consecuencias pueden ser también importantes en genética forense, ya que el hecho de que exista estratificación poblacional puede afectar al cálculo de frecuencias necesario para evaluar el peso de la prueba forense.

Otros aspectos de interés en el estudio del genoma mitocondrial están relacionados con el conocimiento de la filogenia en linajes concretos, lo cual es de vital importancia para el diseño de paneles de SNPs de discriminación poblacional,

o la detección de errores generados *in vitro* (mutaciones fantasma, recombinación artificial, etc), conocimiento de las tasas de mutación, etc.

Este trabajo de tesis se enmarca en este contexto, y para ello se han estudiado varias cohortes de pacientes de cáncer de mama, esquizofrenia, enfermedad meningocócica y autismo para los que existía una hipótesis bien establecida que defendía un papel importante de distintas variantes mitocondriales con la patología o algunos aspectos clínico/fenotípicos como por ejemplo la aparición temprana de la esquizofrenia y el haplogrupo JT. Al mismo tiempo, también nos hemos interesado por profundizar en el conocimiento del linaje más frecuente existente en poblaciones europeas o de ancestralidad europea (R0) así como en alguna aplicación forense como la utilización de mtSNPs para el genotipado de muestras muy degradadas.

## Objetivos generales

- ✓ Análisis de la variabilidad del ADN mitocondrial, mediante el genotipado de determinados polimorfismos dispersos a lo largo del genoma o secuenciación directa, con el objetivo de obtener una visión global de esta variabilidad en las poblaciones europeas y sus posibles aplicaciones biomédicas.
- ✓ Análisis crítico de los estudios de asociación entre enfermedades complejas y variabilidad del ADNmt (haplogrupos mitocondriales).
- ✓ Estudio y revisión de los datos publicados de genomas completos en poblaciones europeas, en particular, el linaje R0 que comprende más del ~60% de la variabilidad en estas poblaciones; con la finalidad de actualizar las filogenias involucradas (y su nomenclatura) para que sirvan de marco de referencia para estudios genético poblacionales y forenses.

## Objetivos específicos

- ✓ Estudio de la presunta asociación entre linajes y SNPs mitocondriales con el cáncer de mama en dos cohortes diferentes de pacientes con sus respectivos controles sanos.



- ✓ Replicar estudios de asociación entre la variabilidad del ADN mitocondrial y el riesgo a padecer esquizofrenia.
- ✓ Investigación de la existencia o ausencia del papel del genoma mitocondrial en el riesgo a padecer infección meningocócica mediante un estudio de asociación caso-control.
- ✓ Analizar la posible asociación entre la variabilidad del ADNmt y el riesgo de padecer autismo, evaluar el potencial papel patogénico de varias mutaciones mitocondriales y determinar si mutaciones realmente patogénicas, causantes de enfermedades mitocondriales, aparecen en la cohorte de autistas.
- ✓ Discusión sobre las inconsistencias detectadas en publicaciones anteriores en esta línea de investigación, tales como la ausencia de control de la estratificación poblacional, diseño adecuado de muestreo (tamaño muestral), réplica de los resultados, escaso conocimiento filogenético del genoma mitocondrial o planteamiento erróneo de hipótesis.
- ✓ Selección de un conjunto de SNPs codificantes del ADNmt que permita la clasificación de muestras pertenecientes al haplogrupo R0 en sus sublinajes más frecuentes y que pueda ser utilizada como herramienta de discriminación forense.
- ✓ Revisión de la nomenclatura de la filogenia del macro-haplogrupo R0, resumen de los conflictos existentes y propuesta de una nomenclatura actualizada y de unos criterios para evitar futuras inconsistencias.
- ✓ Validación de SNPs para la práctica forense, mediante el análisis de muestras con baja concentración de ADN, cabellos sin bulbo y huesos.



---

*MATERIAL Y MÉTODOS*



## 1 Muestras

Para la realización de esta tesis doctoral se han empleado las muestras que se enumeran a continuación. Todas las muestras se han obtenido con consentimiento informado del donante.

Los **estudios de asociación genético-clínicos** realizados han consistido en el genotipado de grupos de casos (pacientes con cáncer de mama, infección meningocócica, esquizofrenia y autismo) y de los respectivos controles. Todas las muestras de ADN han sido extraídas a partir de sangre:

- 464 muestras de ADN de pacientes de cáncer de mama esporádico de población española peninsular (Vega et al. 2009).
- 453 muestras de ADN de controles (mujeres españolas que no presentaban cáncer de mama) de la población de la península (Vega et al. 2009).
- 302 muestras de ADN pertenecientes a casos de cáncer de mama esporádico de la población las islas Canarias (Vega et al. 2009).
- 295 muestras de ADN de mujeres de las islas Canarias que no presentaban cáncer de mama (Vega et al. 2009).
- 307 muestras de ADN de pacientes de infección meningocócica de población peninsular.
- 616 muestras de ADN de controles para infección meningocócica, individuos de población peninsular que no presentaban la enfermedad.
- 454 muestras de ADN pertenecientes a pacientes españoles de esquizofrenia.
- 159 muestras de ADN de pacientes españoles de autismo fundamentalmente recogidas en Cataluña.
- 137 muestras de ADN de controles para autismo (individuos que no presentaban este trastorno) recogidos en Cataluña.

Para el trabajo relacionado con el área de **genética de poblaciones** se han empleado las siguientes muestras de ADN:

- 282 muestras de ADN de individuos sanos de diferentes comarcas de Galicia, obtenidas a partir de muestras de sangre recogidas por el Centro de Transfusión de Galicia.
- 135 muestras de ADN, extraídas a partir de manchas de sangre, pertenecientes a individuos sanos de tres regiones de Cantabria (Valle del Pas, Valle de Liébana y Santander).
- 101 muestras de ADN pertenecientes a individuos nacidos en Cataluña.
- 8 muestras de ADN, extraídas a partir de manchas de sangre, de individuos del País Vasco.

En relación al trabajo de validación **forense**, se han analizado distintos tipos de muestras, pelos, huesos, dientes y diluciones en diferentes concentraciones.

## 2 Extracción de ADN

### 2.1 Extracción de ADN a partir de mancha de sangre: fenol-cloroformo

Para la extracción de ADN de manchas de sangre y saliva se ha utilizado un método de extracción con fenol-cloroformo especialmente diseñado para manchas de fluidos biológicos:

- Cortar 1cm<sup>2</sup> de tejido y resuspenderlo en 500μL de tampón DLB (Tris-HCl 1M, NaCl 5M, EDTA 0'5M pH 8.0).
- Añadir 50μL de SDS al 10% y 5μL de proteinasa K (20 mg/mL).
- Incubar toda la noche a 56°C con agitación suave.
- Añadir 20μL de NaCl 5M y 575μL de fenol: cloroformo: alcohol isoamílico (25:24:1); mezclar por inversión.
- Centrifugar 3 minutos a 12000 r.p.m. y pasar la fase acuosa a otro tubo.
- Añadir 575μL de cloroformo: alcohol isoamílico y mezclar por inversión.
- Centrifugar 3 minutos a 12000 r.p.m.
- Pasar la fase acuosa a otro tubo y añadir 1mL de etanol absoluto frío.

- Mantener durante 15 minutos a  $-80^{\circ}\text{C}$  y centrifugar 15 minutos a 12000 r.p.m.
- Vaciar el contenido de los tubos y dejar secar el precipitado al aire.
- Resuspender el precipitado en 50-100 $\mu\text{L}$  de agua bidestilada estéril e incubar a  $56^{\circ}\text{C}$  con agitación suave entre 2 y 16 horas.
- El rendimiento final es de aproximadamente 35ng/ $\mu\text{L}$ .

## 2.2 Extracción de ADN a partir de sangre

La extracción de ADN de las muestras empleadas en los estudios de asociación que se presentan en este trabajo ha sido realizada en los centros de recogida de las muestras y posteriormente enviadas al Instituto de Medicina Legal de Santiago de Compostela. Los protocolos empleados para la extracción de ADN se describen aquí de forma orientativa.

La extracción de ADN genómico de casos y controles para el estudio de cáncer de mama se llevó a cabo a partir de linfocitos de sangre periférica, empleando extracción automática de ADN (MagNAPure, Roche, Mannheim, Alemania) y siguiendo los protocolos recomendados por el fabricante.

Tanto las muestras de ADN de casos con infección meningocócica como las muestras de los controles de este estudio, fueron extraídas a partir de un volumen de 300 $\mu\text{L}$  de sangre total mediante Wizard® Genomic DNA Purification Kit (Promega, Wisconsin, EE.UU.), siguiendo los protocolos recomendados.

Para la extracción de ADN de casos de esquizofrenia se utilizó el robot Chemagic Magnetic Separation Module I, siguiendo las instrucciones del fabricante (<http://www.chemagen.com>) partiendo de un volumen inicial de 5mL de sangre.

## 2.3 Extracción de ADN a partir de otros vestigios biológicos

Las muestras de ADN forenses de tejido óseo (dos dientes y dos huesos) fueron extraídas mediante un protocolo estándar de fenol-cloroformo del Instituto de Medicina Legal de Santiago de Compostela. Los principales pasos de este protocolo se presentan aquí de forma resumida:

- Limpieza de huesos y piezas dentarias.
- Fragmentación de las epífisis de los huesos y de las piezas dentarias.
- Pulverización de fragmentos de hueso y dientes.
- Digestión del producto de la pulverización.
- Extracción y purificación del ADN.

Para la extracción de ADN a partir de pelos se empleó el BioRobot EZ1 (Qiagen, Hilden, Alemania), usando las condiciones propuestas por el fabricante.

### 3 Genotipado de SNPs de la región codificante mediante minisecuencia

En general, el genotipado de SNPs, dispersos a lo largo del genoma mitocondrial o nuclear, comienza por una PCR multiplex que consiste en amplificar en una misma reacción varias regiones simultáneamente, de forma que cada una de ellas contenga uno o varios de los SNPs de interés. Para que este paso tenga éxito, es necesario diseñar de forma adecuada tanto los *primers* de PCR como los de minisecuencia.

#### 3.1 Diseño de *primers* para una PCR multiplex

Las consideraciones más importantes que se deben tener en cuenta a la hora de diseñar un multiplex se podrían resumir en los siguientes puntos:

- Número moderado de SNPs.
- Tamaño adecuado de los amplicones.
- Temperatura de *melting* ( $T_m$ ) similar entre todos los *primers* y sondas.
- Especificidad de los *primers* para la región de interés.
- Evitar interacciones entre los *primers*.
- Comprobar la ausencia de formación de estructuras secundarias dentro del propio *primer* (horquillas).



- Reducción del número de amplicones, incluyendo más de un SNPs por amplicón, siempre que sea posible.

### 3.1.1 Selección de SNPs

El primer paso en el diseño de un multiplex consiste en la selección de los SNPs de interés. Para ello es esencial tener en cuenta la población de estudio, ya que la informatividad de cada polimorfismo puede variar de una población a otra, y llevar a cabo una revisión bibliográfica previa. En el presente trabajo, se han genotipado SNPs diagnóstico de haplogrupos mitocondriales europeos, así como mutaciones patogénicas confirmadas y posiciones “probablemente” patogénicas.

### 3.1.2 Diseño de primers

Para diseñar los *primers* de amplificación, el empleo de herramientas bioinformáticas es extremadamente útil. En nuestro caso, se han utilizado los siguientes programas:

- *Software Primer3* (<http://frodo.wi.mit.edu/cgi-bin/primer3>)

Este software, de acceso libre, se ha empleado para el diseño de *primers*, y permite fijar parámetros como: tamaño de los *primers* y de los amplicones, rango óptimo para la Tm (60°C, aproximadamente), proporción de G/C de los oligonucleótidos,...

- *BLAST* (<http://www.ncbi.nlm.nih.gov/BLAST/beta/Blast.cgi>)

Este programa gratuito pertenece al *National Center for Biotechnology Information* (NCBI), que contiene una base de datos de secuencias gracias a la cual se puede comprobar la especificidad de los *primers* diseñados, mediante la evaluación del porcentaje de alineación con otras secuencias.

- *Autodimer*

(<http://www.cstl.nist.gov/biotech/strbase/AutoDimerHomepage/AutoDimerProgramHomepage.htm>)

El autodimer es un programa (descarga gratuita) que chequea si existen potenciales estructuras en horquilla y posibles dímeros de *primers* del multiplex. Estableciendo un número máximo de identidades, el programa

reporta una lista de las interacciones que podrían ocurrir y la temperatura a la que tiene lugar dichas interacciones.

- *Oligonucleotide Properties Calculator*

(<http://www.basic.northwestern.edu/biotools/oligocalc.html>)

Este software, de acceso libre, permite comprobar la  $T_m$  de cada *primer* y chequear si presenta complementariedad interna.

### 3.2 Reacción en cadena de la polimerasa (PCR)

En los diferentes trabajos que aquí se presentan se han desarrollado varios multiplexes diferentes, pero todos ellos se optimizaron para las siguientes condiciones: 5ng de muestra de ADN, 3 $\mu$ L de QIAGEN Multiplex PCR Kit (Qiagen), 1 $\mu$ L de *mix* de *primers* y se añade agua hasta un volumen final de 10 $\mu$ L. Las concentraciones de cada pareja de *primers* vienen indicadas en las respectivas tablas. El PCR *master mix* contiene los reactivos necesarios para la PCR (Taq polimerasa, dNTPs a concentraciones óptimas, cofactores como el MgCl<sub>2</sub>, tampón,...) y además está optimizado para la amplificación de multiplexes.

Las reacciones de amplificación se llevaron a cabo en un 9700 Thermocycler (Applied Biosystems, Foster City, CA, EE.UU.). Las condiciones de PCR se resumen en la siguiente tabla:

	Proceso	Temperatura (°C)	Tiempo
<b>Inicio</b>	Activación enzimática	95	15'
	Desnaturalización	94	30"
<b>Ciclos (30)</b>	Anillamiento	58	1'30"
	Extensión	72	1'30"
	Extensión final	72	10'
<b>Fin</b>		4	$\infty$

**Tabla 5:** Condiciones de la reacción de PCR multiplex.

El primer multiplex se basa en el desarrollado en un trabajo previo (Quintáns et al. 2004), con unas modificaciones posteriores, ya que se añadieron dos nuevos marcadores: 13966 y 10463, y en el diseño actual la amplificación de los 19 SNPS se realiza en una sola reacción de PCR.

SNP	Primer forward	Primer reverse	Tamaño (pb)	C.F. (μM)
3010	CAATAACTTGACCAACGGAACA	CGGTCTGAACTCAGATCACGTA	180	0'12
3915/ 3992	TAGCAGAGACCAACCGAACC	GAAGATTGTAGTGGTGAGGGTGT	158	0'32
4216	CTCTACACAACATATTTTGTCAACCAAG	GGTTTGAGGGGGAATGCTGGAG	195	0'2
4336	GGAGCTTAAACCCCTTATTTTC	GATAGGTGGCACGGAGAATTT	80	0'12
4529/ 4580	CAACCCGTCATCTACTCTACCAT	CTTCTGTGGAACGAGGGTTTATT	148	0'2
4769/ 4793	CCGGACAATGAACCATAACC	TGGGTAACCTCTGGGACTCA	118	0'32
6776	GCTTCCTAGGGTTTATCGTGTG	GAGTGTGGCGAGTCAGCTAAA	140	0'12
7028	CACCGTAGGTGGCCTGACTGGC	GTGTAGCCTGAGAATAGGGG	168	0'12
10398/ 10400/ 10463	AAATTGCCCTCCTTTTACCCCTA	TGTAAATGAGGGGCATTTGG	224	0'4
10873	CATAATTTGAATCAACACAACCACC	GTTAGGGGGTCGGAGGAAAAGGTTG	123	0'12
12308 <sup>1</sup>	CTGCTAACTCATGCCCCCATG	ATTACTTTTATTTGGAGTTGCACCAAGATT	106	0'4
12705	TGTAGCATTGTTTCGTTACATGG	AGTTGGAATAGGTTGTTAGCGG	147	0'2
13966	TCACTTTCCTAGGACTTCTAACAGC	GGTTAGGTCTAGGAGGAGTAGGG	191	0'2
14766	TCAACTACAAGAACACCAATGACC	GGAGGTCGATGATGAGTGG	82	0'2

**Tabla 6:** *Primers* de amplificación para el multiplex de 19 SNPs. Las secuencias de los *primers forward* y *reverse* están indicadas en dirección 5'-3'; el tamaño hace referencia al del amplicón resultante; C.F., concentración final de cada pareja de *primers* en la reacción de PCR. 1, los *primers* empleados para la amplificación del SNP 12308 son los diseñados por Torroni et al. (1996). La secuencia del *primer reverse*, tal y como se indica en el trabajo de Quintáns et al. (2004), presenta una modificación en una de las bases cercanas al extremo 3' (subrayada) y que fue incluida para detectar este polimorfismo mediante RFLP.

*A posteriori* surgió la necesidad de analizar otros seis mtSNPs para realizar una réplica de los resultados obtenidos en un estudio contemporáneo (Bai et al. 2007), de forma que se desarrolló un nuevo multiplex.

SNP	<i>Primer Forward</i>	<i>Primer reverse</i>	Tamaño (pb)	C.F. (μM)
<b>3197</b>	GCCTTCCCCGTAAATGATA	TTATGCGATTACCGGGCTCT	98	0'2
<b>8994</b>	AATGCCCTAGCCCACTTCTT	AGGTGGCCTGCAGTAATGTT	140	0'05
<b>9055</b>	CCTACTCATGCACCTAATTGGA	GGCTTACTAGAAGTGTGAAAACGTA	155	0'125
<b>11719</b>	CAGCCATTCTCATCAAACC	GCGTTCGTAGTTTGAGTTTGC	113	0'25
<b>13708</b>	AACGAAAATAACCCACCCCTA	GCGGGGGAAATGTTGTTAGT	95	0'3
<b>16519</b>	GCCCATAACACTTGGGGGTA	GGGAACGTGTGGGCTATTTA	92	0'2

**Tabla 7:** *Primers* de amplificación para el multiplex de 6 SNPs. Las secuencias de los *primers forward* y *reverse* están indicadas en dirección 5'-3'; el tamaño hace referencia al del amplicón resultante; C.F., concentración final de cada pareja de *primers* en la reacción de PCR.

En conjunto, estos dos multiplexes permiten la clasificación de muestras de poblaciones europeas en los haplogrupos mitocondriales más frecuentes. Se han empleado ambos en tres de los estudios de asociación realizados (cáncer de mama, infección meningocócica y esquizofrenia) mientras que para el estudio de autismo sólo se ha genotipado el multiplex de 19 SNPs.

Con la finalidad de cubrir gran parte de la filogenia del macrohaplogrupo R0, muy frecuente en poblaciones europeas, se desarrollaron tres multiplexes en los que se genotiparon 71 SNPs de la región codificante del ADNmt (Álvarez-Iglesias et al. 2009; Mosquera-Miguel et al. 2009); tablas 8, 9 y 10:

SNP	<i>Primer forward</i>	<i>Primer reverse</i>	Tamaño (pb)	C.F. (μM)
709/750	GGCTCACATCACCCCATAAA	CGTTTTGAGCTGCATTGCT	161	0'2
2581/2706	GCCTGCCCAGTGACACAT	GCTCCATAGGGTCTTCTCGT	195	0'2
3010 <sup>1</sup>	CAATAACTTGACCAACGGAACA	CGGTCTGAACTCAGATCACGTA	179	0'4
3796/3847	TCAACATTACTAATAAGTGGCTCCTTT	GGTTCGGTTGGTCTCTGCTA	135	0'2
4550/4580 <sup>1</sup>	CAACCCGTCATCTACTCTACCAT	CTTCTGTGGAACGAGGGTTTATT	148	0'2
6253/6296/ 6365	TGACTCTTACCTCCCTCTCTCC	GATGAAATTGATGGCCCCTA	189	0'2
6776 <sup>1</sup>	GCTTCCTAGGGTTTATCGTGTG	GAGTGTGGCGAGTCAGCTAAA	140	0'4
7337	GGCTCATTCATTTCTCTAACAGC	TCCAGGTTTATGGAGGGTTC	110	0'4
10810	GCTAAAACTAATCGTCCCAACA	AATTAGGCTGTGGGTGGTTG	97	0'2
12858/12957	CAACACAGCAGCCATTCAAG	GAGGCCTAGTAGTGGGGTGA	157	0'2
13708/13759	AACGAAAATAACCCACCCCTA	GTTGTTTGAAGGGGGATG	113	0'2
14365/14470	CCACCCCATCATACTCTTTCA	TAGGGGGAATGATGGTTGTC	159	0'3
14766/14770 <sup>1</sup>	TCAACTACAAGAACCAATGACC	GGAGGTCGATGATGAGTGG	82	0'2
15218	ACTATCCGCCATCCCATACA	GGGCAAGATGAAGTGAAAGG	110	0'4

**Tabla 8:** *Primers* de amplificación para el multiplex 1 de Álvarez-Iglesias et al. (2009). Las secuencias de los *primers forward* y *reverse* están indicadas en dirección 5'-3'; el tamaño hace referencia al del amplicón resultante; C.F., concentración final de cada pareja de *primers* en la reacción de PCR. 1, *Primers* publicados en Quintáns et al. (2004).

SNP	Primer forward	Primer reverse	Tamaño (pb)	C.F. (μM)
951/961	TCACACGATTAACCCAAGTCA	ACTCAGGTGAGTTTGTAGCTTTATTG	87	0'2
3915/3936/ 3992 <sup>1</sup>	TAGCAGAGACCAACCGAACC	GAAGATTGTAGTGGTGAGGGTGT	158	0'6
4310/4336	AGCATTCCCCCTCAAACCTA	TTTGGATTCTCAGGGATGG	127	0'4
4727/4745/ 4769/4793	TCCTTCTAATAGCTATCCTCTTCAACA	TGGGTAACCTCTGGGACTCA	154	0'4
7028 <sup>1</sup>	CACCGTAGGTGGCCTGACTGGC	GTGTAGCCTGAGAATAGGGG	168	0'4
7645	ACATGCAGCGCAAGTAGGTC	AAAATGATTATGAGGGCGTGA	90	0'2
8269/8271	TAGGGCCCGTATTTACCCTAT	AAGAGGTGTTGGTTCTCTTAATCTTT	110	0'2
8473/8592/ 8598/8602	CCCAACTAAAAATATTAAACACAAACT	GGAGGTGGGGATCAATAGAG	193	0'2
9066/9088/ 9150	CCTACTCATGCACCTAATTGGA	GGCTTACTAGAAGTGTGAAAACGTA	155	0'3
10044	CCGTAACTTCCAATTAAGTAGTTTTG	AAGGCTAGGAGGGTGTGATT	91	0'6
10394	CCATGAGCCCTACAAACAACCT	TGAGTCGAAATCATTTCGTTTTG	159	0'3
13404	TATGTGCTCCGGGTCCATC	TGGTGAGGGAGGTTGAAGTG	104	0'2

**Tabla 9:** *Primers* de amplificación para el multiplex 2 de Álvarez-Iglesias et al. (2009). Las secuencias de los *primers forward* y *reverse* están indicadas en dirección 5'-3'; el tamaño hace referencia al del amplicón resultante; C.F., concentración final de cada pareja de *primers* en la reacción de PCR. 1, *Primers* publicados en Quintáns et al. (2004).

SNP	Primer Forward	Primer reverse	Tamaño (pb)	C.F. (μM)
1438	AACCTAAGGGTCGAAGGTGGA	AGGGCCCTGTTCAACTAAGC	66	0'1
2259	TCAAGCTCAACACCCACTACC	TGCGGAGGAGAATGTTTTCA	131	0'2
5250/ 5263	ATTCCATCCACCCTCCTCTC	GGTGGGGATGATGAGGCTAT	111	0'6
8869	GGACTCCTGCCTCACTCATTT	AAGTGGGCTAGGGCATTTTTT	128	0'4
8994	AATGCCCTAGCCCACTTCTT	AGGTGGCCTGCAGTAATGTT	140	0'1
9336	GCCATGTGATTTCACTTCCA	GTGGCCTTGGTATGTGCTTT	117	0'2
10166/ 10211	ACTACCACAACCTCAACGGCTACA	AGGGGTAAAAGGAGGGCAAT	145	0'2
11140	CATTACAGCCACAGAACTAATCAT	GTTCTGGCTGGTTGCCTCAT	99	0'2
11719	CAGCCATTCTCATCCAAACC	GCGTTCGTAGTTTGAGTTTGC	113	0'3
12308 <sup>1</sup>	CTGCTAACTCATGCCCCATG	ATTACTTTTATTTGGAGTTGCACCAAGATT	106	0'3
12438	CCACCCTAACCTGACTTCC	GTGGATGCGACAATGGATTT	106	0'1
12705 <sup>1</sup>	TGTAGCATTGTTTCGTTACATGG	AGTTGGAATAGGTTGTTAGCGG	147	0'2
13101/ 13105	CAGCCCTACTCCACTCAAGC	TGGGCTATTTTCTGCTAGGG	83	0'2
14869/ 14872	CAACATCTCCGCATGATGAA	AGGCGTCTGGTGAGTAGTGC	104	0'2
15452A	AGACGCCCTCGGCTTACTT	GTCGCCTAGGAGGTCTGGTG	78	0'2
15773	CCGCAGACCTCCTCATCTA	CGGATGCTACTTGTCCAATG	81	0'1
15833/ 15904	CCCTTTTACCATCATTTGGACA	AAAGGTTTTTCATCTCCGGTTT	162	0'2

**Tabla 10:** *Primers* de amplificación para el multiplex 3 de Álvarez-Iglesias et al. (2009). Las secuencias de los *primers forward* y *reverse* están indicadas en dirección 5'-3'; el tamaño hace referencia al del amplicón resultante; C.F., concentración final de cada pareja de *primers* en la reacción de PCR. 1, *Primers* publicados en Quintáns et al. (2004).

Con el objetivo de genotipar las principales mutaciones patogénicas confirmadas del genoma mitocondrial, para el estudio de autismo, se empleó el multiplex diseñado en el trabajo de Álvarez-Iglesias et al. (2008) (tabla 11). Además se genotiparon otras seis posiciones descritas como potencialmente patogénicas para este desorden (Weissman et al. 2008), y los *primers* diseñados para su amplificación se muestran en la tabla 12.

Mutación	Primer Forward	Primer reverse	Tamaño (pb)	C.F. (μM)
3243	TATACCCACACCCACCCAAG	GGCCATGGGTATGTTGTTAAG	118	0'2
3460	CCGAACGAAAAATTCTAGGC	GCGGTGATGTAGAGGGTGAT	153	0'2
3697	GCCTAGCCGTTTACTCAATCC	TGAGATTGTTTGGGGCTACTGC	94	0'15
3946/3949 <sup>1</sup>	TAGCAGAGACCAACCGAACC	GAAGATTGTAGTGGTGAGGGTGT	157	0'2
7445	CCCTACCACACATTCTGAAGAA	TGGCTTGAAACCAGCTTTG	89	0'2
8993	AATGCCCTAGCCCACTTCTT	AGGTGGCCTGCAGTAATGTT	140	0'15
9176	AAATCGCTGTGCGCTTAATC	TCATTAGGAGGGCTGAGAGG	154	0'3
10158/10191	TCAACACCCTCCTAGCCTTA	GGGTAAAAGGAGGGCAATTT	196	0'3
10663	ACACCCACTCCCTCTTAGCC	GGCCATATGTGTTGGAGATTG	110	0'3
11777/11778/ 11832	CACGGGCTTACATCCTCATT	GGGGGTAAGGCGAGGTTAG	157	0'2
12706 <sup>1</sup>	TGTAGCATTGTTCTGTTACATGG	AGTTGGAATAGGTTGTTAGCGG	146	0'2
13513/13514	ATTGGCAGCCTAGCATTAGC	CAGGGAGGTAGCGATGAGAG	131	0'2
14459/14482/ 14484/14487	CTCCATCGCTAACCCCACTA	TTCTGAATTTTGGGGGAGGT	170	0'4

**Tabla 11:** *Primers* de amplificación para el multiplex de mutaciones patogénicas confirmadas (Álvarez-Iglesias et al. 2008). Las secuencias de los *primers forward* y *reverse* están indicadas en dirección 5'-3'; el tamaño hace referencia al del amplicón resultante; C.F., concentración final de cada pareja de *primers* en la reacción de PCR. 1, *Primers* publicados en Quintáns et al. (2004).

Posición	Primer Forward	Primer reverse	Tamaño (pb)	C.F. (μM)
3394/3397 <sup>2</sup>	CCGAACGAAAAATTCTAGGC	GCGGTGATGTAGAGGGTGAT	154	0'2
4295	GCATTCCCCCTCAAACCTAA	CTAGAAATAAGGGGGTTTAAGCTCCTAT	89	0'2
10394 <sup>1</sup>	CCATGAGCCCTACAAACAACT	TGAGTCGAAATCATTCGTTTTG	159	0'2
11809 <sup>2</sup>	CACGGGCTTACATCCTCATT	GGGGGTAAGGCGAGGTTAG	158	0'2
11984	AACCACGTTCTCCTGATCAAA	GTGAGTGAGCCCCATTGTGT	111	0'1

**Tabla 12:** *Primers* de amplificación para el multiplex de mutaciones potencialmente patogénicas. Las secuencias de los *primers forward* y *reverse* están indicadas en dirección 5'-3'; el tamaño hace referencia al del amplicón resultante; C.F., concentración final de cada pareja de *primers* en la reacción de PCR. 1, *Primers* publicados en Álvarez-Iglesias et al. (2009); 2, *Primers* publicados en Álvarez-Iglesias et al. (2008).



### 3.3 Purificación enzimática del producto de PCR: ExoSAP-IT®

Los productos de PCR, una vez chequeados en un gel de acrilamida y antes de la reacción de minisecuencia, deben ser tratados con ExoSAP-IT® (USB® Corporation, Cleaveland, Ohio, EE.UU.). ExoSAP-IT® utiliza dos enzimas hidrolíticas, Exonucleasa I, que elimina los *primers* monocatenarios residuales y cualquier ADN de cadena simple, y fosfatasa alcalina de camarón (*Shrimp Alkaline Phosphatase*), que se encarga de desfosforilar los dNTPs no incorporados durante la reacción de amplificación, impidiendo así que interfieran en reacciones posteriores.

Para ello se incubaba 1 µL de producto de PCR con 0.5 µL de enzima con el siguiente programa:

	Temperatura (°C)	Tiempo
<b>Incubación</b>	37	15'
<b>Inactivación</b>	85	15'
	4	∞

**Tabla 13:** Condiciones para la purificación enzimática del producto de PCR mediante ExoSAP-IT®.

### 3.4 Reacción de minisecuencia

La reacción de minisecuencia se lleva a cabo con el Kit SNaPshot™ (AB), que contiene un enzima termoestable y sin actividad exonucleasa 3'-5' (AmpliTaQ® FS) y cuatro ddNTPs marcados fluorescentemente:

ddNTP	Marcador	Color
A	d6G	VERDE
C	dTAMRA™	NEGRO
G	dR110	AZUL
T	dROX™	ROJO

**Tabla 14:** Marcaje empleado en el SNaPshot™ Kit (AB)

A la hora de diseñar las sondas, el objetivo es distribuir las en el electroferograma en función de su tamaño y de los alelos posibles para cada SNP, y para ello hay que tener en cuenta las siguientes consideraciones:

- Las sondas deben abarcar la zona adyacente al SNP, ya sea corriente arriba o corriente abajo del mismo.
- Especificidad con el ADN molde, aunque con menores restricciones que para el diseño de *primers* de PCR.
- Evitar interacciones entre sondas.
- Las sondas deben presentar tamaños distintos entre ellas, de forma que en la electroforesis no se solapen y se puedan diferenciar bien los alelos de cada sistema. Para ello, se modifica la longitud de las sondas incorporando colas no homólogas de nucleótidos en el extremo 5'. Esto permite incluir un número elevado de marcadores en una sola reacción.

El tamaño de la sonda, añadiéndole el nucleótido incorporado, define la posición del marcador y el color del ddNTP determina el alelo del SNP.

La reacción de minisequenciación tiene lugar en un volumen final que puede variar entre 5 y 10  $\mu\text{L}$ , en función del número de SNPs a genotipar en una misma reacción, y que comprende 1-3  $\mu\text{L}$  de SNaPshot™ kit (AB), 1'5  $\mu\text{L}$  de producto de PCR purificado, 0'5-1  $\mu\text{L}$  de *mix* de sondas, completando con agua hasta el volumen final. La reacción se lleva a cabo en un 9700 Thermocycler (AB), siguiendo las recomendaciones del fabricante.

	Proceso	Temperatura (°C)	Tiempo
25 ciclos	Desnaturalización	96	10''
	Anillamiento	50	5'
	Extensión	60	30''
Fin		4	$\infty$

**Tabla 15:** Condiciones para la reacción de minisequenciación con el kit SNaPshot™.

Las sondas o *primers* de minisequenciación empleados en los diferentes trabajos se resumen en las siguientes tablas donde, entre otros parámetros, se indica la longitud de cada sonda y los alelos del polimorfismo (rCRS-variante, respectivamente).

Los dos multiplexes con los que se llega a una clasificación de las poblaciones europeas presentan también dos reacciones de minisecuencia independiente, de 19 y 6 sondas, respectivamente. Los tamaños de dichos *primers* de extensión varían entre 20 y 64pb.

SNP	Primer de minisecuencia	Tamaño (pb)	Alelo	Cadena	C.F. (μM)
14766	AATGACCCCAATACGCAAAA	20	C-T	L	0'2
4529	CTTTGCAGGCACACTCATCAC	21	A-T	L	0'16
3010	GGAGATGTTGGATCAGGACATCCC	24	G-A	L	0'2
7028	TACACGACACGTACTACGTTGTAGC	25	C-T	L	0'16
4580	TTACCTGAGTAGGCCTAGAAATAAACAT	28	G-A	L	0'2
3915	(tgac) <sub>2</sub> TAAGCCTGAGACTAGTTCGGACTC	32	G-A	H	0'36
4216	(c) <sub>4</sub> TACCACTCACCTAGCATTACTTATATGA	33	T-C	L	0'2
10400	(c) <sub>10</sub> CGTTTTGTTTAACTATATACCAATTC	37	C-T	H	0'4
4769	(gact) <sub>3</sub> GGGCTATTCCTAGTTTTATTGCTATAGC	40	A-G	H	0'4
12308	(c) <sub>22</sub> CATTGGTCTTAGGCCCAA	41	A-G	L	0'4
3992	t(gact) <sub>5</sub> CCCTATTCTTCATAGCCGAATACA	45	C-T	L	0'36
10873	(c) <sub>15</sub> GTTGTTGTTGATTGTTGTTAAAAATAGTAG	45	T-C	H	0'2
12705	(c) <sub>17</sub> AACATTAATCAGTTCTTCAAATATCTACTCAT	49	C-T	L	0'2
4793	(gact) <sub>7</sub> CTACTCAGAAGTGAAAGGGGGC	50	A-G	H	0'4
4336	ct(gact) <sub>7</sub> GCTTAAACCCCTTATTTCTAGGAC	55	T-C	L	0'6
6776	act(gact) <sub>5</sub> CGTGTGTCTACGTCTATTCCTACTGTAAATAT	55	T-C	H	0'2
10398	(gact) <sub>7</sub> ATGAGTGACTACAAAAGGATTAGACTGA	58	A-G	L	0'4
10463	ct(gact) <sub>9</sub> ATGTAAATGAGGGGCATTTGGTA	61	T-C	H	0'6
13966	act(gact) <sub>9</sub> gaCAATCCCTATCTAGGCCTTCTT	64	A-G	L	0'2

**Tabla 16:** *Primers* de minisecuencia para el multiplex de 19 SNPs modificado a partir de dos multiplexes de un trabajo anterior (Quintáns et al. 2004). Las secuencias de las sondas están indicadas en dirección 5'-3', las regiones en minúsculas se corresponden con las colas no homólogas; la cadena indica sobre qué hebra del ADNmt ocurre el anillamiento del *primer*; C.F., concentración final de la sonda en la reacción de minisecuencia.

SNP	Primer de minisecuencia	Tamaño (pb)	Alelo	Cadena	C.F. (μM)
9055	CTACTCATGCACCTAATTGGAAGC	24	G-A	L	0'125
3197	CCCCGTAAATGATATCATCTCAACTTAG	28	T-C	L	0'2
16519	ct(gact) <sub>2</sub> GACATCTGGTTCCTACTTCAGGG	33	T-C	L	0'3
13708	(gact) <sub>4</sub> CTACTAAACCCCATTAAGGCCTG	40	G-A	L	0'3
11719	ct(gact) <sub>7</sub> ATTCTCATAATCGCCACGG	50	G-A	L	0'5
8994	act(gact) <sub>9</sub> GCCTACTCATTCAACCAATAGCCCT	64	G-A	L	0'075

**Tabla 17:** *Primers* de minisecuencia para el multiplex de seis SNPs. Las secuencias de las sondas están indicadas en dirección 5'-3', las regiones en minúsculas se corresponden con las colas no homólogas; la cadena indica sobre qué hebra del ADNmt ocurre el anillamiento del *primer*; C.F., concentración final de la sonda en la reacción de minisecuencia.

La minisecuencia de los 71 SNPs pertenecientes al haplogrupo R0 se llevó a cabo en tres reacciones, con las sondas que se describen en las siguientes tablas:

SNP	Primer de minisecuencia	Tamaño (pb)	Alelo	Cadena	C.F. (μM)
2581	TGATTATGCTACCTTTGCACGGT	23	A-G	H	0'2
750	CTCTAAATCACCACGATCAAAAGG	24	A-G	L	0'2
709	(gact) <sub>3</sub> TTACACATGCAAGCATCCCC	32	G-A	L	0'2
2706	t(gact) <sub>2</sub> AGGGTCTTCTCGTCTTGCTGTGT	32	A-G	H	0'2
3010	t(gact) <sub>3</sub> AACCTTTAATAGCGGCTGCACCAT	37	G-A	H	0'2
3796	ct(gact) <sub>2</sub> CTAATAAGTGGCTCCTTTAACCTCTCC	37	A-G	L	0'1
3847	(gact) <sub>5</sub> ATTACTCCTGCCATCATGACCC	42	T-C	L	0'2
4580	ct(gact) <sub>3</sub> TTACCTGAGTAGGCCTAGAAATAAACAT	42	G-A	L	0'2
6253	ct(gact) <sub>6</sub> TGTTCTGCTCCGGCCTCCACT	48	T-C	H	0'1
6296	(gact) <sub>6</sub> AACAGGTTGAACAGTCTACCCTCC	48	C-T	L	0'2
6365	ct(gact) <sub>7</sub> GATGGCCCCTAAGATAGAGGAGAC	54	T-C	H	0'2
10810	act(gact) <sub>5</sub> CAACAATTATATTACTACCATTGACATGACT	54	T-C	L	0'2
6776	ct(gact) <sub>6</sub> CGTGTGTCTACGTCTATTCTACTGTAAATAT	58	T-C	H	0'3
12858	act(gact) <sub>8</sub> GCAGCCATTCAAGCAATCCTATA	58	C-T	L	0'3
7337	(gact) <sub>10</sub> TGATTTGAGAAGCCTTCGCTTC	62	G-A	L	0'3
12957	(gact) <sub>9</sub> CAACAAATAGCCCTTCTAAACGCTAA	62	T-C	L	0'3
13759	(gact) <sub>10</sub> TTCTCATTACTAACAACATTTCCCCC	66	G-A	L	0'1
14766	t(gact) <sub>11</sub> CAATGACCCCAATACGCAAAA	66	C-T	L	0'3
14365	ct(gact) <sub>11</sub> GTTAGCGATGGAGGTAGGATTGGT	70	C-T	H	0'3
15218	(gact) <sub>11</sub> TCCTCAGATTTCATTGAACTAGGTCTG	70	A-G	H	0'3
14770	(gact) <sub>11</sub> GAATGAGTGGTTAATTAATTTTATTAGGGG	74	C-T	H	0'3
4550	ct(gact) <sub>13</sub> GCGCTAAGCTCGCACTGATT	74	T-C	L	0'3
14470	t(gact) <sub>13</sub> CCTCAATAGCCATCGCTGTAGTATA	78	T-A/C	L	0'4
13708	ct(gact) <sub>13</sub> CTACTAAACCCATTAAAGGCTG	78	G-A	L	0'3

**Tabla 18:** *Primers* de minisecuencia para el multiplex 1 de Álvarez-Iglesias et al. (2009). Las secuencias de las sondas están indicadas en dirección 5'-3', las regiones en minúsculas se corresponden con las colas no homólogas; la cadena indica sobre qué hebra del ADNmt ocurre el anillamiento del *primer*; C.F., concentración final de la sonda en la reacción de minisecuencia.

SNP	Primer de minisequenciación	Tamaño	Alelo	Cadena	C.F. (μM)
951	CTTTATTGGGGAGGGGGTGAT	21	G-A	H	0'2
4336	AGGGATGGGTTCGATTCTCAT	21	T-C	H	0'2
3915 <sup>1</sup>	(gact) <sub>2</sub> GAAGCCTGAGACTAGTTCGGACTC	32	G-A	H	0'2
4745	gactAATGAACCATAACCAATACTACCAATCA	32	A-G	L	0'2
3992	ct(gact) <sub>3</sub> CCCTATTCTTCATAGCCGAATACA	38	C-T	L	0'3
4310	TCTGATAAAAAGAGTTACTTTTGATAGAGTAAATAATAGG	38	A-G	L	0'4
4727	ct(gact) <sub>5</sub> TACTCTCCGGACAATGAACCAT	44	A-G	L	0'2
7028	act(gact) <sub>4</sub> TACACGACACGTACTACGTTGTAGC	44	C-T	L	0'2
4769	ct(gact) <sub>4</sub> ACCAATCAATACTCATCATTAATAATCATAAT	50	A-G	L	0'2
7645	(gact) <sub>6</sub> GCTACTTCCCCTATCATAGGAGAGCT	50	T-C	L	0'2
4793	t(gact) <sub>5</sub> ATAATCATAATAGCTATAGCAATAAACTAGGAAT	56	A-G	L	0'4
8473	ct(gact) <sub>6</sub> AAAAATATTAAACACAACTACCACCTACC	56	T-C	L	0'2
8269	(gact) <sub>8</sub> gacTGAAATAGGGCCCGTATTTACCCTATA	62	G-A	L	0'06
9066	ct(gact) <sub>8</sub> GTGTAGAGGGAAGGTTAATGGTTGATAT	62	A-G	H	0'3
8592	act(gact) <sub>11</sub> GCCTACCCGCCGCAGTACT	66	G-A	L	0'1
10044	(gact) <sub>9</sub> TTAAGGCGAAGTTTATTACTCTTTTTTGAA	66	A-G	H	0'6
9150	t(gact) <sub>12</sub> CCTAGAAATCGCTGTCGCCTT	70	A-G	L	0'2
10394	ct(gact) <sub>9</sub> CTGGCCTATGAGTGACTACAAAAAGGATTAGA	70	C-T	L	0'3
8602	(gact) <sub>13</sub> CGCCGCAGTACTGATCATTCTA	74	T-C	L	0'4
9088	(gact) <sub>10</sub> AGTAGAATTAGAATTGTGAAGATGATAAGTGTAG	74	T-C	H	0'2
13404	ct(gact) <sub>11</sub> ATCATCCACAACCTTAACAATGAACAAGATAT	78	T-C	L	0'3
3936	t(gact) <sub>14</sub> TGCGGCGTATTCGATGTTGAA	78	C-T	H	0'6
8271	(gact) <sub>15</sub> GGGCCCCGTATTTACCCTATAGC	82	A-T	L	0'3
8598	act(gact) <sub>15</sub> ACCCGCCGCAGTACTGATCAT	84	T-C	L	0'4
961	(gact) <sub>15</sub> CGTAAAGAGTGTTTGTAGATCACCCCC	86	T-G	L	0'06

**Tabla 19:** *Primers* de minisequenciación para el multiplex 2 de Álvarez-Iglesias et al. (2009). Las secuencias de las sondas están indicadas en dirección 5'-3', las regiones en minúsculas se corresponden con las colas no homólogas; la cadena indica sobre qué hebra del ADNmt ocurre el anillamiento del *primer*; C.F., concentración final de la sonda en la reacción de minisequenciación. <sup>1</sup>, *Primers* de minisequenciación publicados en Quintáns et al. (2004).

SNP	Primer de minisecuencia	Tamaño (pb)	Alelo	Cadena	C.F. (μM)
2259	TCCCAAACATATAACTGAACTCCTCA	26	C-T	L	0'2
5250	TTCTTCGATAATGGCCCATTTGGGCA	26	T-C	H	0'2
5263	act(gact) <sub>2</sub> GGCTTTTGGCCAAATGGG	30	C-T	L	0'2
8869	t(gact) <sub>2</sub> TTATGAGCGGGCACAGTGATT	30	A-G	L	0'6
9336	act(gact) <sub>3</sub> TTCCACTCCATAACGCTCCTC	36	A-G	L	0'15
10211	(gact) <sub>4</sub> gactgactgactTCCCCGCGCGTCCCTTT	36	C-T	L	0'15
10166	ct(gact) <sub>4</sub> GGCTACATAGAAAAATCCACCCC	41	T-C	L	0'3
12308 <sup>1</sup>	(c) <sub>22</sub> CATTGGTCTTAGGCCCCAA	41	A-G	L	0'4
13105	ct(gact) <sub>4</sub> CACTCAAGCACTATAGTTGTAGCAGGA	45	A-G	L	0'2
12705	ct(gact) <sub>3</sub> AACATTAATCAGTTCTTCAAATATCTACTCAT	46	C-T	L	0'2
11719	ct(gact) <sub>7</sub> ATTCTCATAATCGCCACGG	50	G-A	L	0'4
12438	t(gact) <sub>6</sub> CTAACAAAAAACTCATACCCCCA	50	T-C	L	0'15
13101	ct(gact) <sub>7</sub> GAAGCGGATGAGTAAGAAGATTCC	54	A-C	H	0'2
15452	t(gact) <sub>8</sub> GCCCTCGGCTTACTTCTCTTC	54	C-A	L	0'4
15773	act(gact) <sub>9</sub> CCTGAATCGGAGGACAACCA	59	G-A	L	0'1
8994	act(gact) <sub>9</sub> GCCTACTCATTCAACCAATAGCCCT	64	G-A	L	0'06
11140	act(gact) <sub>7</sub> CATATTTTATATCTTCTTCGAAACCACACTTAT	64	C-T	L	0'2
1438	ct(gact) <sub>10</sub> GTCGAAGGTGGATTAGCAGTAAACT	68	A-G	L	0'06
14872	(gact) <sub>12</sub> CTCCTTGGCGCCTGCCTGAT	68	C-T	L	0'4
15904	(gact) <sub>10</sub> TTTCATCTCCGGTTTACAAGACTGGTGTATTA	72	C-T	H	0'1
15833	act(gact) <sub>9</sub> GACAAGTAGCATCCGTACTATACTTCACAACAATC	74	C-T	L	0'2
14869	gac(gact) <sub>13</sub> GCTCACTCCTTGGCGCCTGCCT	77	G-A	L	0'2

**Tabla 20:** *Primers* de minisecuencia para el multiplex 3 de Álvarez-Iglesias et al. (2009). Las secuencias de las sondas están indicadas en dirección 5'-3', las regiones en minúsculas se corresponden con las colas no homólogas; la cadena indica sobre qué hebra del ADNmt ocurre el anillamiento del *primer*; C.F., concentración final de la sonda en la reacción de minisecuencia. <sup>1</sup>, *Primers* de minisecuencia publicados en Quintáns et al. (2004).

Las sondas de minisecuencia diseñadas para el multiplex de mutaciones patogénicas confirmadas (Álvarez-Iglesias et al. 2008) y para las seis posiciones adicionales potencialmente patogénicas se muestran en las siguientes tablas. Para genotipar las posiciones 3949, 14484 y 14487 fue necesaria la utilización de

*primers* degenerados debido a la posibilidad de co-existencia de dos mutaciones adyacentes, lo que provocaría dificultades en anillamiento de la sonda.

Mutación	Primer de minisequenciación <sup>1</sup>	Tamaño (pb)	Alelos	Cadena	C.F. (μM)
10158	ACAACTCAACGGCTACATAGAAAAA	25	T-C	L	0'2
3946	cGAACTAGTCTCAGGCTTCAACATC	25	G-A	L	0'2
10663	gactgCAATATTGTGCCTATTGCCATACTAG	31	T-C	L	0'2
3460	(gact) <sub>2</sub> gGCTACTACAACCCTTCGCTGAC	31	G-A	L	0'2
10191	(gact) <sub>4</sub> AGTGCGGCTTCGACCCTATA	36	T-C	L	0'3
14459	(gact) <sub>2</sub> gacCTCAGGATACTCCTCAATAGCCATC	36	G-A	L	0'2
14484	(gact) <sub>3</sub> ATCGCTGTAGTATATCCAAAGACAACVA	40	T-C	L	0'5
3243	(gact) <sub>4</sub> gaACAGGGTTTGTTAAGATGGCAG	40	A-G	L	0'2
11777	(gact) <sub>5</sub> gCAAACTACGAACGCACTCACAGT	44	C-A	L	0'2
7445	(gact) <sub>5</sub> TCGAAGAACCCGTATACATAAAATCTAG	48	A-G/C	L	0'2
12706	(gact) <sub>9</sub> GCGGTAACCTAAGATTAGTATGGTAATTAGGA	52	T-C	H	0'2
3949	(gact) <sub>7</sub> gaAGTCTCAGGCTTCAACATC <b>R</b> AA	52	T-C	L	0'2
14487	(gact) <sub>7</sub> Gctgtagtatatccaaagacaaccay <b>C</b> A	56	T-C	L	0'6
13513	(gact) <sub>8</sub> TTCTCTCACAGGTTTCTACTCCAAA	56	G-A	L	0'1
13514	(gact) <sub>10</sub> gTGCGGTTTCGATGATGTGG	60	A-G	H	0'2
3697	(gact) <sub>9</sub> ctAAACTCAAACTACGCCCTGATC	60	G-A	L	0'2
11778	(gact) <sub>9</sub> gaGAAGTCCTTGAGAGAGGATTATGATG	64	G-A	H	0'3
11832	(gact) <sub>8</sub> gacTCAAACTCTACTCCCACTAATAGCTTTTT	64	G-A	L	0'1
9176	(gact) <sub>10</sub> gATCCAAGCCTACGTTTTCACACTTC	70	T-C/G	L	0'2
14482	(gact) <sub>10</sub> GCACATCGCTGTAGTATATCCAAAGACAAC	73	C-A/G	L	0'6
8993	(gact) <sub>12</sub> gaCCTACTCATTTCAACCAATAGCCC	76	T-G/C	L	0'2

**Tabla 21:** *Primers* de minisequenciación para el multiplex de mutaciones patogénicas confirmadas de Álvarez-Iglesias et al. (2008). Las secuencias de las sondas están indicadas en dirección 5'-3', las regiones en minúsculas se corresponden con las colas no homólogas; la cadena indica sobre qué hebra del ADNmt ocurre el anillamiento del *primer*; C.F., concentración final de la sonda en la reacción de minisequenciación. Para las sondas correspondientes a las mutaciones 3949, 14484 y 14487 se utilizaron *primers* degenerados que se señalan siguiendo el código IUB (posiciones marcadas en negrita).



Posición	Primer de minisecuenciación	Tamaño (pb)	Alelo	Cadena	C.F. (μM)
3394	TTACCGAACGAAAAATTCTAGGC	23	T-C	L	0'2
3397	(gact) <sub>3</sub> GGGGCCTTTGCGTAGTTGTA	32	A-G	H	0'2
4295	ct(gact) <sub>2</sub> GGGGGTTTAAGCTCCTATTATTTACTC	37	A-G	H	0'2
10394	Ct(gact) <sub>2</sub> CTGGCCTATGAGTGACTACAAAAAGGATTAGA	42	C-T	L	0'2
11809	(gact) <sub>5</sub> CATAATCCTCTCTCAAGGACTTCAAAC	47	T-C	L	0'2
11984	(gact) <sub>7</sub> ACTAGTCACAGCCCTATACTCCCTC	53	T-C	L	0'2

**Tabla 22:** *Primers* de minisecuenciación para el multiplex 3 de Álvarez-Iglesias et al. (2009). Las secuencias de las sondas están indicadas en dirección 5'-3', las regiones en minúsculas se corresponden con las colas no homólogas; la cadena indica sobre qué hebra del ADNmt ocurre el anillamiento del *primer*; C.F., concentración final de la sonda en la reacción de minisecuenciación.

### 3.5 Purificación enzimática del producto de minisecuenciación: SAP

Durante este proceso se eliminan los grupos fosfato de los extremos 5' del ADN y de los ddNTPs no incorporados en la reacción de minisecuenciación para evitar interferencias en la electroforesis. Esto se consigue mediante el tratamiento de los productos de minisecuenciación con SAP (USB®), una fosfatasa alcalina de camarón (*Shrimp Alkaline Phosphatase*) termolábil y con actividad altamente específica. El volumen final de la reacción anterior (5-10 μL) es tratado con 0'5-1 μL de SAP.

	Temperatura (°C)	Tiempo
<b>Incubación</b>	37	60'
<b>Inactivación</b>	85	15'
	4	∞

**Tabla 23:** Condiciones para la purificación enzimática del producto de minisecuenciación mediante SAP.

### 3.6 Electroforesis capilar y análisis de resultados

La preparación de los productos de minisecuenciación purificados para la electroforesis capilar implica la adición a 2 μL de dicho producto de 9'5 μL de formamida (HiDi™ Formamide, AB) y 0'3 μL de marcador interno de peso

molecular (GeneScan™-120 LIZ® Size Standard, AB). La formamida se emplea para desnaturalizar las muestras de ADN y el marcador interno de peso molecular, que consiste en nueve fragmentos de tamaño conocido (15 / 20 / 25 / 35 / 50 / 62 / 80 / 110 / 120pb) marcados fluorescentemente, permite inferir el tamaño de los productos de minisequenciación.

La electroforesis capilar tuvo lugar en dos tipos de secuenciadores automáticos: ABI PRISM 3130xl Genetic Analyzer (AB) y ABI PRISM 3730xl Genetic Analyzer (AB), utilizando las condiciones electroforéticas que vienen por defecto en cada secuenciador. En el caso del primer secuenciador, se utilizaron indistintamente los polímeros POP6® y POP7®, mientras que para el ABI 3730 sólo se empleó POP7®.

Los resultados fueron analizados con el programa GeneMapper™ v3.2 Software (AB) que permite una lectura automatizada mediante la creación de paneles específicos para cada multiplex. Para la conversión del formato resultante de este programa en un formato más manejable, se empleó el programa SNPator ([http://bioinformatica.cegen.upf.es/public/new\\_login/index.php](http://bioinformatica.cegen.upf.es/public/new_login/index.php)). Todos los electroferogramas fueron revisados manualmente para comprobar la correcta asignación de los alelos. Además, gracias a las características del genoma mitocondrial, los resultados pudieron ser analizados desde un punto de vista filogenético, permitiendo la detección de inconsistencias. En aquellos casos en los que el resultado obtenido difería de lo esperado, se llevó a cabo una confirmación mediante secuenciación del amplicón en ambas direcciones.

Paralelamente al genotipado de las muestras, se realizaron los mismos análisis con los controles negativos de extracción y PCR, para comprobar la ausencia de contaminación a lo largo de todo el proceso.

## **4 Secuenciación de la región control del ADNmt**

### **4.1 Diseño de primers**

El diseño de los *primers* para amplificar la región control no difiere de lo mencionado anteriormente. Al igual que para los *primers* de la región codificante,

en este caso también es conveniente que la  $T_m$  de los diferentes oligonucleótidos sea similar ya que en ocasiones es preciso combinarlos.

## 4.2 Amplificación de la región control

En esta ocasión el objetivo de la PCR es amplificar un solo fragmento de la región control, aunque el tamaño del amplicón puede variar entre 200 y 1200pb en función de la combinación de *primers* seleccionada.

El número de *primers* diseñados para amplificar la región control es considerablemente alto ya que las muestras de baja calidad precisan ser amplificadas en fragmentos solapantes de pequeño tamaño. La nomenclatura de los *primers* consta de una parte numérica, que indica la posición 3' terminal del primer, mientras que la letra hace referencia a la cadena ligera (L) o la pesada (H) con la que se corresponde la secuencia del primer. Las secuencias de los *primers* empleados se muestran en la siguiente tabla:

<i>Primer</i>	Secuencia 5'-3'
15971L	TTAACTCCACCATTAGCACC
15997L <sup>1</sup>	CACCATTAGCACCCAAAGCT
16159L	TACTTGACCACCTGTAGTAC
16380L	TCAGATAGGGGTCCCTTGAC
16236H	CTTTGGAGTTGCAGTTGATG
16401H <sup>2</sup>	TGATTTACGGAGGATGGTG
16555L	CCCACACGTTCCCCTTAAAT
017H	CCCGTGAGTGGTTAATAGGGT
072H	AGGATGAGGCAGGAATCAAA
172L	ATTATTTATCGCACCTACGT
332L	CCCGCTTCTGGCCACAGCAC
370L	CCCTAACACCAGCCTAACCA
285H	GGGGTTTGGTGGAAATTTTTTG
408H <sup>2</sup>	CTGTTAAAAGTGCATACCGCCA
599H <sup>3</sup>	TTGAGGAGGTAAGCTACATA
611H	AGTGTATTGCTTTGAGGAGG

**Tabla 24:** Listado de *primers* empleados para la amplificación y secuenciación de la región control del genoma mitocondrial. Algunos de los *primers* fueron tomados de publicaciones previas: (1)(Ward et al. 1991); (2) (Vigilant et al. 1989); (3) (Brandstätter et al. 2004).

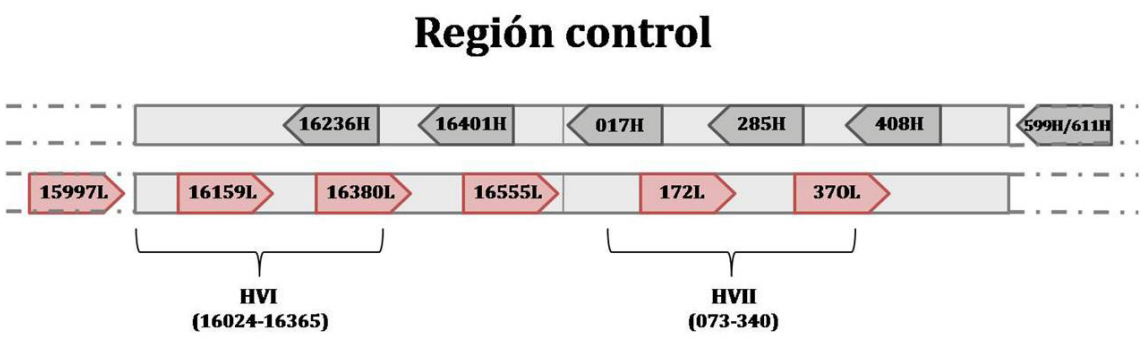


Figura 9: Esquema de los *primers* empleados principalmente para amplificar la región control del ADNmt.

A alrededor de 400 de las muestras analizadas con el multiplex de R0 se les genotipó el rango 16024-16365 de la región control, mientras que para las restantes (~120 muestras) se secuenció desde 16024 hasta 16569. Para algunas muestras además se procedió al genotipado del resto de la región control (1-576). En 210 muestras pertenecientes al estudio caso-control de autismo se genotipó la región control entera, mientras que para 86 muestras se secuenció desde 16024 hasta 16569. También se secuenció este rango en el caso de las muestras forenses.

La reacción de PCR se llevó a cabo en un 9700 Thermocycler (AB) y las condiciones empleadas son las siguientes:

Reactivo	Concentración
<i>Primer forward</i>	0'3µM
<i>Primer reverse</i>	0'3µM
BSA	0'16µg/µL
<i>Buffer</i>	1X
MgCl <sub>2</sub>	1'5mM
dNTPs	200µM
Taq polimerasa	2'5U
ADN	5ng
Agua	Hasta 25µL

**Tabla 25:** Protocolo para la reacción de amplificación de la región control del ADNmt.

Dependiendo de la naturaleza y calidad de las muestras, las condiciones de PCR pueden haber sufrido modificaciones, principalmente en lo que se refiere al número de ciclos y a la temperatura de *melting* (T<sub>m</sub>), dentro de los rangos que se muestran en la siguiente tabla:

Proceso		Temperatura (°C)	Tiempo
<b>Inicio</b>		95	1'
<b>Ciclos (30-36)</b>	Desnaturalización	95	10"
	Anillamiento/ <i>melting</i>	55-60	30"
	Extensión	72	30"
<b>Fin</b>	Extensión final	72	10'
		4	∞

**Tabla 26:** Condiciones para la reacción de amplificación de la región control del ADNmt.

### 4.3 Purificación post-PCR

#### 4.3.1 Purificación post-PCR para secuenciación con dRhodamine

En la purificación de los productos de PCR que fueron secuenciados mediante dRhodamine se emplearon columnas de MICROSPIN™ HR S-300 (Pharmacia Biotech, Uppsala, Sweden) como se indica en Salas et al. (1998).

#### 4.3.2 Purificación post-PCR para secuenciación con Big-Dye®

La purificación del producto de PCR, en la mayoría de los casos, se llevó a cabo mediante filtración en placas MultiScreen® PCR µ96 (Millipore, Billerica, Massachusetts, EE.UU.), de manera que los *primers* y dNTPs no incorporados son eliminados, mientras que los fragmentos amplificados permanecen en la membrana de la placa. Para ello, se resuspende todo el producto de PCR en agua bidestilada hasta completar un volumen final de 100µL, se transfiere a la placa y se filtra. Tras la filtración, el producto retenido en la membrana se recupera mediante resuspensión en 25-30µL de agua bidestilada e incubación durante 5-10 minutos.

Alternativamente, se puede purificar el producto de PCR mediante digestión enzimática con Exo-SAP-IT® (USB®), para lo cual se incuban 3µL de producto de PCR con 1'5µL de enzima con el programa indicado en la tabla 13, pero alargando el tiempo de incubación a 30 minutos.

### 4.4 Reacción de secuenciación

Para la secuenciación de las muestras se emplearon dos kits diferentes: dRhodamine Terminator Cycle Sequencing Kit (AB) y BigDye® Terminator v3.1

Cycle Sequencing Kit (AB), ambos basados en el método de secuenciación de Sanger. Los dos kits incluyen una ADN polimerasa termoestable modificada, sin actividad exonucleasa 3'-5', dNTPs, y ddNTPs marcados con cuatro fluorocromos distintos.

#### 4.4.1 Secuenciación con dRhodamine

El volumen final de la reacción es de 5 $\mu$ L, resultado de la incorporación de 2'5 $\mu$ L de producto de PCR purificado, 2 $\mu$ L de kit de secuenciación y 3'2 pmoles de *primer*. El programa utilizado en el termociclador fue:

	Proceso	Temperatura (°C)	Tiempo
<b>Inicio</b>		95	15'
<b>Ciclos (25-30)</b>	Desnaturalización	95	15"
	Anillamiento	55	30"
	Extensión	70	30"
<b>Fin</b>	Extensión final	72	10'
		4	$\infty$

**Tabla 27:** Condiciones para la reacción de secuenciación con dRhodamine de la región control del ADNmt.

#### 4.4.2 Secuenciación con BigDye®

En un volumen final de 10 $\mu$ L se incorporaron 3-4 $\mu$ L de producto de PCR purificado, 0'5 $\mu$ L de kit, 2 $\mu$ L de *buffer* 5X, 3pmoles del *primer* y agua hasta completar el volumen final.

	Proceso	Temperatura (°C)	Tiempo
<b>Inicio</b>		96	1'
<b>Ciclos (25-30)</b>	Desnaturalización	96	10"
	Anillamiento	50	5"
	Extensión	60	4'
<b>Fin</b>		4	$\infty$

**Tabla 28:** Condiciones para la reacción de secuenciación con BigDye® de la región control del ADNmt.

### 4.5 Purificación post-secuenciación

El objetivo de este paso es eliminar los ddNTPs no incorporados durante la reacción de secuencia y así evitar su migración durante la electroforesis.

#### 4.5.1 Purificación alcohólica tras secuenciación con dRhodamine

- Añadir a las muestras 3'75µL de MgCl<sub>2</sub> y 13'75µL de etanol absoluto.
- Centrifugar durante 20 minutos a 13000 r.p.m.
- Eliminar el etanol con una punta de pipeta.
- Añadir 64'75µL de etanol 70% y centrifugar 15 minutos a 13000 r.p.m.
- Eliminar primero el etanol por decantación y a continuación con una punta de pipeta.
- Dejar evaporar por completo el etanol antes de resuspender en formamida para cargar la muestra en el secuenciador.

#### 4.5.2 Purificación tras secuenciación con BigDye®

El primer paso consiste en una purificación enzimática, similar a la empleada en minisequenciación, en la que se añade 1µL de SAP al volumen final de la reacción de secuenciación (10µL). El programa es el mismo que el indicado en la tabla 23, pudiendo alargar el tiempo de incubación a 80 minutos.

A continuación se lleva a cabo una purificación mediante filtración en placa Montage™ SEQ<sub>96</sub> Sequencing Reaction Cleanup Kit (Millipore):

- Resuspender con 25µL de la solución de inyección (EDTA 0'3mM, pH= 8) el volumen de la reacción de secuenciación y transferir todo el volumen a la placa de filtración.
- Filtrar hasta que todo el volumen desaparezca y añadir de nuevo 25µL de la misma solución.
- Filtrar y añadir 40µL de solución para hacer un último lavado.
- Añadir 20-30µL de solución a cada pocillo e incubar a temperatura ambiente durante 10 minutos con agitación suave.
- Resuspender suavemente y recuperar todo el volumen transfiriéndolo a la placa que irá a continuación al secuenciador.

De forma alternativa y tras la purificación enzimática, puede realizarse una purificación alcohólica (normalmente realizada en placa):

- Para un volumen de 11 $\mu$ L de producto de secuenciación purificado enzimáticamente, añadir 1'1 $\mu$ L de EDTA 125mM, 1'1 $\mu$ L de NaOAc 3M y 27'5 $\mu$ L de etanol absoluto.
- Dar un vórtex a la placa y centrifugar 45 minutos a 3700 r.p.m.
- Dar la vuelta a la placa rápidamente sobre un papel de filtro y centrifugar un minuto a 200 r.p.m.
- Añadir 38'5 $\mu$ L de etanol 70% y centrifugar 15 minutos a 3700 r.p.m.
- Dar la vuelta a la placa rápidamente sobre un papel de filtro y centrifugar un minuto a 200 r.p.m.
- Introducir la placa abierta en la estufa durante 5-10 minutos y cuando esté completamente seca, resuspender en 11 $\mu$ L de formamida.

#### **4.6 Electroforesis capilar y análisis de resultados**

Las secuencias que fueron tratadas con una purificación alcohólica fueron resuspendidas en formamida HiDi™ (AB) mientras que aquellas secuenciadas con BigDye® (AB) y purificadas mediante filtración en placa se resuspendieron en solución de inyección. La electroforesis capilar tuvo lugar en tres secuenciadores: ABI PRISM 3100® Genetic Analyzer (AB), ABI PRISM 3130xl® Genetic Analyzer (AB), y ABI PRISM 3730xl® Genetic Analyzer (AB) con las condiciones de los módulos por defecto de cada secuenciador.

Las secuencias resultantes fueron analizadas con el programa Sequencing Analysis 5.3.1 (AB), y alineadas con la Secuencia de Referencia de Cambridge revisada (rCRS) (Andrews et al. 1999), utilizando el programa SeqScape 2.5 (AB).

### **5 Secuenciación de genomas mitocondriales completos**

#### **5.1 Amplificación de fragmentos solapantes**

Para la obtención de los diferentes fragmentos solapantes mediante los que se amplifica todo el genoma mitocondrial, se emplearon 5ng de muestra de ADN, 4 $\mu$ L de QIAGEN Multiplex PCR Kit (Qiagen), 5pmoles de cada *primer*, y se completa con agua hasta un volumen final de 10 $\mu$ L.



La amplificación se llevó a cabo en un 9700 Thermocycler (AB). Las condiciones de PCR, que se resumen en la siguiente tabla:

	Proceso	Temperatura (°C)	Tiempo
<b>Inicio</b>	Activación enzimática	95	15'
	Desnaturalización	94	30"
<b>Ciclos (35)</b>	Anillamiento	58	1'30"
	Extensión	72	1'30"
	Extensión final	72	10'
<b>Fin</b>		4	∞

**Tabla 29:** Condiciones para la reacción de PCR mediante la que se amplifica el genoma mitocondrial completo en 11 fragmentos solapantes.

Se emplearon los *primers* desarrollados por Torroni (2001) para amplificar todo el genoma mitocondrial humano en 11 fragmentos solapantes con un tamaño medio de 1800 bases aproximadamente. En la siguiente tabla se muestra el tamaño de cada amplicón resultante y las secuencias de los *primers* de PCR empleados:

Amplicón	Tamaño amplicón (pb)	Primers de amplificación			
		Nombre	Secuencia (5'→3')	Posición 5'	Posición 3'
<b>1</b>	1.845	14897F	CTAGCCATGCACTACTCACCAG	14897	14918
		155R	AATAGGATGAGGCAGGAATCAA	155	134
<b>2</b>	1.759	16488F	CTGTATCCGACATCTGGTTCCT	16488	16509
		1677R	GTTTAGCTCAGAGCGGTCAAGT	1677	1656
<b>3</b>	1.832	1404F	ACTTAAGGGTCTGAAGGTGGATT	1404	1425
		3235R	CTTAACAAACCCTGTTCTTGGG	3235	3214
<b>4</b>	1.784	2900F	CAATAACTTGACCAACGGAACA	2900	2921
		4683R	TTAGAAGGATTATGGATGCGGT	4683	4662
<b>5</b>	1.771	4381F	ACCTATCACACCCCATCTAAA	4381	4402
		6151R	ACTAGTCAGTTGCCAAAGCCTC	6151	6130
<b>6</b>	1.747	5871F	GCTTCACTCAGCCATTTTACCT	5871	5892
		7617R	TCTTGTAGACCTACTTGCGCTG	7617	7596
<b>7</b>	1.980	7239F	GCATACACCACATGAAACATCC	7239	7260
		9218R	TTGGTGGGTGCTTATGTGTTGT	9218	9197
<b>8</b>	1.740	8910F	CTTACCACAAGGCACACCTACA	8910	8931
		10649R	AGGCACAATATTGGCTAAGAGG	10649	10628
<b>9</b>	1.769	10457F	TCATATTTACCAATGCCCCTC	10457	10478
		12225R	AGTTCTTGTGAGCTTTCTCGGT	12225	12204
<b>10</b>	1.816	12014F	CTCACCACACATTAACAACA	12014	12035
		13829R	AGTCCTAGGAAAGTGACAGCGA	13829	13808
<b>11</b>	1.873	13477F	GCAGGAATACCTTTCCTCACAG	13477	13498
		15349R	GTGCAAGAATAGGAGGTGGAGT	15349	15328

**Tabla 30:** Listado de los *primers* empleados en la amplificación del genoma mitocondrial en 11 fragmentos solapantes (Torroni et al. 2001).

## 5.2 Purificación del producto de PCR en gel de agarosa

El objetivo de este paso es permitir la extracción del fragmento de tamaño deseado, eliminando *primers*, dNTPs y otras impurezas que pueden interferir en reacciones posteriores.

Para la preparación del gel de agarosa se pesan 0'8g de agarosa y se diluyen en 100mL de tampón TBE 1X (Tris/Borato/EDTA, tampón de electroforesis) aplicando calor, y se añaden 4μL de bromuro de etidio. Tras la electroforesis, se extrae la banda de interés empleando una pipeta *Pasteur*. La agarosa con la banda extraída del gel será introducida en tubos *ependorf* de 1'5mL para continuar con su purificación. Para esta purificación se utiliza el QIAquick Gel Extraction Kit (Qiagen) según las recomendaciones del fabricante.

## 5.3 Reacción de secuenciación

La secuenciación se llevó a cabo empleando BigDye® Terminator v3.1 Cycle Sequencing Kit (AB).

Se añaden 3μL de producto de PCR purificado, 0'5μL de kit, 2'5μL de *buffer* 5X, 10pmoles del *primer* y agua hasta completar un volumen final de 12'5μL. La secuenciación se llevó a cabo en un 9700 Thermocycler (AB):

	Proceso	Temperatura (°C)	Tiempo
<b>Inicio</b>		96	3'
<b>Ciclos (25)</b>	Desnaturalización	96	30"
	Anillamiento	50	15"
	Extensión	60	4'
<b>Fin</b>		4	∞

**Tabla 31:** Condiciones para la reacción de secuenciación de fragmentos grandes del ADNmt con BigDye®.

Para la secuenciación de cada amplicón se emplean en general tres *primers forward* (excepto para el amplicón seis), y las condiciones se detallan en la siguiente tabla (Torróni et al. 2001):

Amplicón	Nombre	Primers de secuenciación		
		Secuencia (5'→3')	Posición	Posición
			nucleotídica 5'	nucleotídica 3'
1	14948F	CACATCACTCGAGACGTAAA	14948	14967
	15564F	ATTTCTTATTGCGCTACACA	15564	15583
	131R	ACAGATACTGCGACATAGGG	131	112
	16522F	TAAAGCCTAAATAGCCCACA	16522	16541
2	584F	TAGCTTACCTCCTCAAAGCA	584	603
	1060F	AAGACCCAAACTGGGATTAG	1060	1079
	1445F	GAGTGCTTAGTTGAACAGGG	1445	1464
3	2047F	TTTAAATTTGCCACAGAAC	2047	2066
	2509F	ATCACCTCTAGCATCACCAG	2509	2528
	3085F	ATCCAGGTCGGTTTCTATCT	3085	3104
4	3598F	CTCAACCTAGGCCTCCTATT	3598	3617
	4010F	ACACCCTCACCCTACAATC	4010	4029
	4410F	CAGCTAAATAAGCTATCGGG	4410	4429
5	5014F	CCTCAATTACCCACATAGGA	5014	5033
	5544F	TCAAAGCCCTCAGTAAGTTG	5544	5563
6	6041F	CCTTCTAGGTAACGACCACA	6041	6060
	6600F	CACCTATTCTGATTTTTTCGG	6600	6619
	7336F	CGAAGCGAAAAGTCCTAATA	7336	7355
7	7937F	TTCAACTCCTACATACTTCCC	7937	7957
	8459F	AACTACCACCTACCTCCCTC	8459	8478
	8975F	TCATTCAACCAATAGCCC	8975	8992
8	9589F	AAGTCCCACTCCTAAACACA	9589	9608
	10147F	ACATAGAAAAATCCACCCCT	10147	10166
	10498F	TAGCATTTACCATCTCACTTCT	10498	10519
9	11081F	ATAACATTACAGCCACAGA	11081	11100
	11644F	CCTCGTAGTAACAGCCATTC	11644	11663
	12114F	ACATCATTACCGGGTTTTTC	12114	12132
10	12600F	ATTCATCCCTGTAGCATTGT	12600	12619
	13134F	AGCAGAAAAATAGCCCACTAA	13134	12153
	13568F	TTACTCTCATCGCTACCTCC	13568	13587
11	14103F	CTCTTTCTTCTTCCCACTCA	14103	14122
	14603F	GAAGGCTTAGAAGAAAACCC	14603	14622

**Tabla 32:** Listado de los *primers* empleados en la secuenciación del genoma mitocondrial (Torroni et al. 2001).

## 5.4 Purificación post-secuenciación y electroforesis capilar

Este paso es idéntico a los explicados en la sección 4.5.2 y 4.6, aunque de forma alternativa se puede llevar a cabo la purificación con Montage™ SEQ<sub>96</sub> Sequencing Reaction Cleanup Kit (Millipore) seguido de una purificación con Sephadex™ G-10 (Amersham Biosciences, Uppsala, Suecia), empleando las recomendaciones del fabricante.

## 6 Análisis estadísticos

### 6.1 Contraste de hipótesis: Test $\chi^2$

En el contraste de hipótesis, la hipótesis nula ( $H_0$ ) es que no existe asociación entre el SNP y el estatus de la enfermedad, mientras que la hipótesis alternativa ( $H_1$ ) asume que sí hay asociación.

La finalidad del test  $\chi^2$  es comprobar si existe independencia entre la variable estatus de la enfermedad (caso-control) y la presencia o ausencia de una determinada variante genética. Ello se realiza comparando lo “observado” (frecuencias muestrales) frente lo que cabría esperar en caso de que efectivamente fuesen independientes. Obviamente, diferencias entre lo observado y lo esperado existirán siempre, pero el test  $\chi^2$  permite discernir si dichas diferencias pueden ser atribuidas al azar o si efectivamente existe asociación entre las variables. El p-valor obtenido para un test  $\chi^2$  nos indica la probabilidad de que los resultados observados se deban únicamente al azar, de ahí que consideremos que existe asociación cuando un p-valor es pequeño (en muchas ocasiones,  $\alpha=0.05$  es considerado el “umbral de decisión”).

Para este análisis se utilizaron programas informáticos que crean tablas de contingencia 2 x 2.

<i>SNP1</i>	Caso	Control	
Alelo rCRS	a	b	a + b
Variante	c	d	c + d
	a + c	b + d	Total

**Tabla 33:** Tabla de contingencia para los valores observados.

<i>SNP1</i>	Caso	Control
Alelo rCRS	$\frac{(a + b) \times (a + c)}{\text{Total}}$	$\frac{(a + b) \times (b + d)}{\text{Total}}$
Variante	$\frac{(c + d) \times (a + c)}{\text{Total}}$	$\frac{(c + d) \times (b + d)}{\text{Total}}$

**Tabla 34:** Tabla de contingencia para el cálculo de los valores esperados.

Con estos datos se calcula el valor del estadístico  $\chi^2$ , que es la suma de los cuadrados de la diferencia absoluta entre los valores observados y esperados en relación al valor esperado:

$$\chi^2 = \sum \frac{(O - E)^2}{E}$$

El valor de  $\chi^2$  puede ser comparado con valores tabulados cuando el tamaño muestral es muy grande. Existen tablas con estos valores en función del umbral de decisión seleccionado ( $\alpha$ ) y de los grados de libertad. Los grados de libertad son igual al producto del número de filas menos 1, (r-1), por el número de columnas menos 1, (k-1). Así, para el caso en el que se estudie la relación entre dos variables dicotómicas (tabla 2x2) los g.l. son 1. El test  $\chi^2$  es un test no dirigido (test de planteamiento bilateral), que indica si existe o no relación entre dos factores pero no en qué sentido se produce tal asociación.

De ser cierta la hipótesis nula, el valor obtenido debería estar dentro del rango de mayor probabilidad según la distribución  $\chi^2$  correspondiente. El p-valor que usualmente reportan la mayoría de paquetes estadísticos es la probabilidad de obtener, según esa distribución, un dato más extremo que el que proporciona el test. Si el p-valor es muy pequeño (usualmente se considera  $p < 0.05$ ) es poco probable que se cumpla la hipótesis nula y se debería de rechazar.

Cuando los valores observados sean iguales o inferiores a cinco, se emplea el test exacto de Fisher.

## 6.2 Cálculo de Odds Ratio (OR)

El Odds Ratio (OR) es el cociente entre la probabilidad de que un evento suceda entre la probabilidad de que no suceda. Es una medida que cuantifica el efecto de la variable genética sobre el estatus de la enfermedad.

$$OR = \frac{a \times d}{b \times c} \quad I.C. = OR \times \exp \left( \pm 1.96 \times \sqrt{\frac{1}{a} + \frac{1}{b} + \frac{1}{c} + \frac{1}{d}} \right)$$

Siempre toma valores positivos:

- El valor neutro es uno: no asociación
- Valores mayores de uno indican factor de riesgo

- Valores menores de uno indican factor protector

El I.C. permite cuantificar la significancia del valor de OR obtenido. Sin embargo, habitualmente, sólo se tiene en cuenta un OR estadísticamente significativo (es decir, aquel cuyo intervalo de confianza no contiene el uno) si éste va acompañado de un valor de P significativo para un test de contraste de hipótesis.

### 6.3 Corrección por test múltiple

Cuando realizamos un test de asociación (como el  $\chi^2$ ) existe una probabilidad (igual a 0'05 si hemos considerado ese “umbral de decisión”) de cometer un error cuando consideremos que existe asociación entre una variante y la enfermedad. El problema es que cuando se estudian varios SNPs la probabilidad de que alguno de ellos aparezca como estadísticamente asociado, para el umbral de decisión adoptado y debido al azar, aumenta de forma proporcional al número de SNPs o hipótesis estudiadas. La forma de solucionarlo es, generalmente, restringiendo el “umbral de decisión”, que debe ajustarse en función del número de hipótesis planteadas. Existen varios métodos para corregir el umbral de significación en estos casos, entre ellos:

- **Corrección Bonferroni:** corrige el umbral de decisión dividiéndolo por el número de test que se están llevando a cabo. Es excesivamente conservativo y, sin embargo, la corrección universalmente aceptada.
- **Test de permutación test:** consiste en obtener N p-valores permutando la variable estatus de la enfermedad en casos y controles, es decir, obtener una distribución empírica de los valores de  $\chi^2$  cuando cambiamos las etiquetas de casos y de controles a los individuos de nuestra muestra, pero manteniendo los genotipos. El p-valor corregido será igual a la proporción de los N p-valores menores que el p-valor original.

## 6.4 Análisis de supervivencia: log Rank

Se denomina análisis de supervivencia al conjunto de técnicas que permiten estudiar la variable “tiempo hasta que ocurre un evento” y su dependencia de otras posibles variables explicativas, que a su vez pueden ser fijas en el tiempo (sexo, marcadores genéticos,...) o variables. El tiempo de supervivencia se define como el tiempo transcurrido desde el acontecimiento o estado inicial hasta el estado final.

En nuestro caso, el objetivo era analizar si existían diferencias significativas en las curvas de supervivencia de pacientes de esquizofrenia en función de sus linajes mitocondriales.

Para comparar si las diferencias observadas en dos curvas de supervivencia pueden ser explicadas o no por el azar, se empleó la prueba log Rank, que compara el número de eventos observados en cada grupo con el número de eventos esperados. El número esperado de pérdidas para un grupo se calcula multiplicando el número total de pérdidas en un periodo dado por la proporción de pacientes en ese grupo. Se emplea la prueba  $\chi^2$  para analizar si existen diferencias entre las pérdidas observadas y esperadas, siendo la hipótesis nula que las distribuciones de supervivencia son iguales en ambos grupos.

## 6.5 Regresión logística

Los modelos de regresión logística son modelos estadísticos en los que se desea conocer la relación entre una variable dependiente cualitativa dicotómica (regresión logística binaria o binomial) o con más de dos valores (regresión logística multinomial) y una o más variables explicativas independientes (covariables), ya sean cualitativas (dicotómicas) o cuantitativas. La ecuación logística de partida es:

$$P(y = 1|x) = \frac{e^{b_0 + \sum_{i=1}^n b_i x_i}}{1 + e^{b_0 + \sum_{i=1}^n b_i x_i}}$$

Siendo  $P(y=1|x)$  la probabilidad de que  $y$  tome el valor uno en presencia de las covariables  $x$  ( $x_1, x_2, \dots, x_n$ ).  $b_0$  es la constante del modelo o el término independiente,  $n$  es el número de covariables,  $b_i$  los coeficientes de las covariables

y  $x_i$  las covariables que forman parte del modelo. La ecuación inicial del modelo es de tipo exponencial, aunque su transformación logarítmica (logit) permite su uso como una función lineal:

$$\log\left(\frac{P(y = 1|x)}{1 - P(y = 1|x)}\right) = b_0 + \sum_{i=1}^n b_i x_i$$

Los modelos de regresión logística tienen dos finalidades:

- Explicativa: cuantifica la importancia de la relación existente entre cada una de las covariables y la variable explicativa, teniendo en cuenta la presencia de otros factores.
- Predictiva: construye un modelo que permite predecir el valor de la variable dependiente (clasifica individuos dentro de las categorías de esta variable) dada la presencia de determinadas covariables. También permite analizar el peso a cada una de las variables independientes como predictoras de la variable dependiente.

En nuestro caso, la variable explicativa es dicotómica y hace referencia al estatus de la enfermedad (0 en caso de ausencia de la enfermedad y 1 para presencia).

## 6.6 Análisis filogenéticos

Las estimas de diversidad genética, como la divergencia de secuencia que presentan los genomas de dos especies o individuos, se obtienen mediante la comparación de dichas secuencias. Para la obtención de las estimas de divergencia de secuencia ( $\delta$ ) inter e intrapoblacionales se puede emplear, entre otros, el método Nei y Tajima de máxima verosimilitud, que considera la proporción de posiciones compartidas entre dos haplotipos en relación al número total de bases. Sin embargo, los métodos basados en parsimonia son más comunes en el análisis de secuencias mitocondriales. Así por ejemplo, se pueden emplear algoritmos tales como “Reduced-Median” o “Median-Joining”, que asumen la existencia de reconstrucciones filogenéticas alternativas igualmente parsimoniosas que pueden explicar los mismos estados de carácter. Las relaciones filogenéticas entre un



grupo de organismos pueden representarse gráficamente mediante árboles filogenéticos que consisten en nodos (unidades taxonómicas) conectados por ramas. Para poder distinguir entre un carácter ancestral y uno derivado es necesario que este tipo de árbol tenga una raíz o “*outgroup*” (grupos externos). El árbol elegido es el que requiere el mínimo número de eventos mutacionales para explicar la distribución de los caracteres entre los individuos a estudiar. En ocasiones existen ambigüedades en los datos que se resuelven mediante la formación de redes para indicar los caminos filogenéticos alternativos. Además, se pueden obtener estimas sobre la escala temporal de eventos evolutivos aplicando la tasa de mutación del genoma mitocondrial. Por ejemplo, a partir de la tasa de divergencia de secuencia de dos linajes y empleando una tasa de substitución adecuada, se puede estimar el tiempo que hace desde que ambos linajes divergieron, o lo que es lo mismo, calcular la edad del antepasado común más reciente (TMRCA) de ambos.

## 6.7 Otros análisis realizados

De forma puntual se han utilizado otros procedimientos que se describen brevemente a continuación:

### 1.1.1 Análisis de ancestralidad

El programa Structure (<http://pritch.bsd.uchicago.edu/structure.html>) permite, mediante un modelo basado en *clusters*, inferir la estructura poblacional de una muestra y asignar para cada individuo una probabilidad de pertenencia a diferentes *clusters* (*membership*).

En nuestro caso se empleó un panel de 34 SNPs informativos de ancestralidad (AIMs, *Ancetral Informative Markers*) (Phillips et al. 2007) que se caracterizan por presentar distribuciones de frecuencias alélicas muy contrapuestas entre los principales grupos poblacionales, africanos, nativo americanos y europeos. De esta forma, gracias al genotipado previo de muestras pertenecientes a estos grupos poblacionales, se pudo estimar la ancestralidad de nuestras muestras, intentando controlar así el efecto de estratificación poblacional.

### **1.1.2 Análisis de componentes principales: paquete estadístico R**

El objetivo de este análisis es obtener combinaciones lineales de las variables originales de forma que expliquen la mayor cantidad de variabilidad de la población de estudio. De esta forma, trabajando con los dos (2D) o tres (3D) componentes principales, se pueden obtener gráficos que permitan una fácil interpretación de los datos. En el contexto de genética de poblaciones, se busca dar con combinaciones lineales de los marcadores genéticos que ofrezcan la mayor información posible acerca de las poblaciones en estudio. Se puede conocer además el porcentaje de variabilidad explicado por cada una de las componentes principales.

En este sentido, se llevó a cabo este análisis para evaluar si la variabilidad de los AIMs permitía separar de forma adecuada los principales grupos poblacionales y si nuestros pacientes pertenecían a un único grupo (europeos) o existía estratificación poblacional en nuestra muestra.

### **1.1.3 Mapas de interpolación: SURFER**

Para las representaciones geográficas de las frecuencias de los haplogrupos, se empleó el programa Surfer 8.0. Se incluyeron 24 muestras poblacionales diferentes para representar espacialmente (geográficamente) el patrón de frecuencias observado. La significancia estadística de estos patrones se puso de manifiesto utilizando el análisis de autocorrelación.

### **1.1.4 Análisis de autocorrelación: SAAP**

La autocorrelación espacial puede definirse como la relación entre los valores de una variable en función de la distribución de los puntos del espacio donde esa variable adquiere un valor determinado. En nuestro caso, las variables son las frecuencias de los haplogrupos y las unidades espaciales son las coordenadas que definen la posición espacial (geográfica) de las poblaciones. Se construyeron matrices de conectividad entre las poblaciones con las coordenadas de las mismas. En definitiva, se trata de analizar si existe alguna relación entre la frecuencia de los haplogrupos mitocondriales detectada en las poblaciones analizadas y la distancia geográfica entre las poblaciones.

Los autocorrelogramas, es decir, las representaciones gráficas de los valores del índice de autocorrelación espacial, se calcularon para 10 haplogrupos. Se empleó el programa Spatial Autocorrelation Analysis Program (SAAP) para la detección y evaluación estadística de señales de gradientes o clinas.



---

*RESULTADOS*



## Artículo 1: Is Mitochondrial DNA Variation Associated with Sporadic Breast Cancer Risk?

Ana Mosquera-Miguel, Vanesa Álvarez-Iglesias, Ana Vega, Roger Milne, Antonio Cabrera de León, Javier Benitez, Ángel Carracedo, Antonio Salas

(2008) *Cancer Research* 68:(2): 623-625

### ¿Está la variación del ADN mitocondrial asociada con el riesgo a padecer cáncer de mama esporádico?

#### ABSTRACT/RESUMEN:

Recientemente, Bai et al., (2007) han defendido la asociación de dos haplogrupos mitocondriales, U y K, con el cáncer de mama esporádico. Este estudio presenta sin embargo inconsistencias claras de tipo estadístico y filogenético. Además, es posible que los resultados de asociación positiva de Bai et al. (2007), y de varios estudios previos, hayan sido potenciados por el efecto de la estratificación poblacional. El objetivo principal del presente trabajo ha sido replicar estos resultados de asociación entre los principales linajes mitocondriales y el cáncer de mama esporádico. Para ello se analizaron dos poblaciones españolas independientes a las que se les genotipó una batería de mtSNPs que identifican los principales haplogrupos europeos. Los tamaños muestrales, 464 casos-453 controles en la población peninsular y 302 casos-295 controles en la población de las islas Canarias, garantizan un poder de detección *a priori* de la asociación superior al 80%; en todo caso, el tamaño muestral es muy superior a otros trabajos publicados. No se ha detectado asociación entre el cáncer de mama esporádico y haplogrupos mitocondriales en ninguna de las dos cohortes analizadas después de corregir por test múltiple (permutación). Los resultados confirman las sospechas que existían en relación al papel que desempeña la estratificación poblacional en el incremento de la tasa de error tipo I en los estudios de asociación. El ADNmt es altamente sensible a los problemas de estratificación debido fundamentalmente a su reducido tamaño efectivo de población.









## Artículo 2: New Population and Phylogenetic Features of the Internal Variation within Mitochondrial DNA Macro-Haplogroup R0

Vanesa Álvarez-Iglesias, **Ana Mosquera-Miguel**, María Cerezo, Beatriz Quintáns, María Teresa Zarrabeitia, Ivon Cuscó, María Victoria Lareu, Óscar García, Luis Pérez-Jurado, Ángel Carracedo, Antonio Salas

(2009) *PLoS ONE* 4(4)

### Nuevas características poblacionales y filogenéticas de la variación interna dentro del macro-haplogrupo R0 del ADN mitocondrial

#### ABSTRACT/RESUMEN:

**Contexto:** El macro-haplogrupo R0 engloba al linaje de ADN mitocondrial (ADNmt) más abundante en Eurasia occidental, el haplogrupo H (~40%). Los sub-linajes de R0 están mal caracterizados en la región control y por ello es necesario el análisis de polimorfismos diagnósticos de la región codificante para ganar la resolución en estudios poblacionales y clínicos.

**Metodología/Resultados principales:** En este trabajo se secuenció la región hipervariable I (HVS-I) de 518 individuos de diferentes regiones del norte de Iberia. Los ADNmt pertenecientes a R0 (57%) se genotiparon con un set de 71 SNPs de la región codificante que caracterizan los principales sub-linajes, además de otras ramas menores, de R0. Se encontraron niveles moderados de estratificación poblacional en el norte de la Península Ibérica; por ejemplo, el haplogrupo V alcanza la máxima frecuencia en Cantabria (norte-centro de Iberia), mientras que las frecuencias son menores en Galicia (noroeste de Iberia) y Cataluña (noreste de Iberia). Cuando se compararon con otras poblaciones europeas y de Oriente Medio, se observó que los haplogrupos H1, H3 y H5a mostraban picos de frecuencia en la región Franco-Cantábrica, disminuyendo hacia este y sur de Europa. Además, se ha caracterizado un nuevo clado del haplogrupo H autóctono del País Vasco, H2a5, mediante secuenciación de genomas completos.

Su edad de coalescencia, 15668 años de antigüedad, data del periodo inmediatamente posterior al Último Máximo Glacial.

**Conclusiones/Importancia:** En contraste con otros sub-linajes del haplogrupo H que experimentaron una re-expansión fuera del refugio Franco-Cantábrico después de la última glaciación (por ejemplo, H1 y H3), H2a5 probablemente permaneció confinado a esta área hasta la actualidad.





















**Erratum**

There was a mistake in the computation of the age of haplogroup H2a5. It is only  $1.285 \pm 161$  years old according to the calibration rate used in the manuscript. Its recent age and assuming a local episode of isolation might easily explain its confined location to the Basque Country. This sub-clade has therefore nothing to do with the post-glacial expansions affecting e.g. haplogroup V.

We would like to thank Vicente Cabrera for kindly let us know about this mistake.

### **Artículo 3: Testing the performance of mtSNP minisequencing in forensic samples**

**Mosquera-Miguel A, Álvarez-Iglesias V, Cerezo M, Lareu MV, Carracedo Á, Salas A**

*(2009) Forensic Sci Int Genetics 3: 261-264*

#### **Explorando la técnica de la minisequenciación de SNPs del ADN mitocondrial para el análisis de muestras forenses**

##### **ABSTRACT/RESUMEN:**

Hay un interés creciente en el desarrollo de protocolos eficientes para el genotipado de SNPs (mtSNPs) de la región codificante del ADN mitocondrial (ADNmt). La técnica de minisequenciación se está convirtiendo en un método popular para el genotipado de SNPs, pero su empleo todavía está limitado a unos pocos laboratorios. Esto se debe en parte a la carencia de estudios que examinen su eficiencia y reproducibilidad cuando se emplea en muestras forenses complejas. En este trabajo se evalúa un ensayo de minisequenciación que consiste en 71 mtSNPs diagnóstico de ramas y sub-ramas conocidas de la filogenia de R0, en donde se encuentra agrupado el haplogrupo H, en muestras forenses reales que incluyen huesos y dientes degradados, pelos y diluciones seriadas. El pequeño tamaño de los amplicones de PCR, que varía entre 66 y 195pb, junto con la eficiencia natural de la metodología de minisequenciación, permiten que este multiplex funcione correctamente en todas las muestras analizadas, tanto en muestras con cantidad limitada de ADN como con alto grado de degradación. No se observaron inconsistencias filogenéticas en ninguno de los haplotipos de mtSNPs obtenidos que puedan indicar potenciales artefactos del genotipado debido, por ejemplo, a contaminación o amplificación espuria de pseudogenes mitocondriales en el ADN nuclear (NUMTs).













## Artículo 4: Investigating the Role of Mitochondrial Haplogroups in Genetic Predisposition to Meningococcal Disease

A Salas, L Fachal, S Marcos-Alonso, A Vega, F Martinón-Torres, JM Martinón Sánchez, F Pardo Sánchez, **A Mosquera-Miguel**, M Fondevila, V Álvarez-Iglesias, JM Sánchez Granados, M Silveira Cancela, JL García Rodríguez, S Rey García, C Calvo Monge, V Pineda Solas, C Pérez Caballero Macarrón, JA Alonso Martín, D Arjona Villanueva, P Azcón González de Aguilar, X Allué Martínez, A Concha Torre, JL Ruibal Francisco, M González-Ripoll Garzón, F Giménez Sánchez, A Reparaz Romero, M Ortiz Pallares, ML Millán Millares, MC Martínez Padilla, R Gómez Zafra, Á Castellanos Ortega, J Casado Flores, A Ferrer Barba, A Hurtado Doce, A Bustinza Arriortua, ML Navarro Gómez, E Esteban Torné, S Cañadas Palazón, C Calvo Macías, E Morteruel Arizkuren, J López Bayón, S Martínez Regueira, JL Pérez Navero, MT Alonso Salas, MM Ballesteros García, JI Muñoz Bonet, S Pantoja Rosso, E Bernaola Iturbe, C Goñi Oraven, F Alvarado Ortega, N Molini Menchón, S Fernández Cebrián, M Baca Cots, MF Martins Prata, E Ocete Hita, M Sánchez Fernández, I Quintela Fernández, R López Peña.

(2009) *PLoS ONE* 4(12)

### Evaluando el papel que juegan los haplogrupos mitocondriales como factores genéticos de predisposición a la enfermedad meningocócica

#### ABSTRACT/RESUMEN:

**Contexto y objetivos:** La enfermedad meningocócica constituye una de las principales causas de muerte por infección en los países industrializados. La gran diversidad en la presentación clínica y en el pronóstico de las infecciones causadas por *Neisseria meningitidis* son el resultado de interacciones complejas entre el ambiente y el perfil genético del hospedador. En este trabajo se evaluó si el acervo genético mitocondrial confiere susceptibilidad a la enfermedad meningocócica (EM).

**Metodología/Resultados principales:** Se llevó a cabo un estudio prospectivo a través de una red nacional de investigación en EM (red ESIGEM, Estudio Sobre la Influencia Genética en la Enfermedad Meningocócica, <http://www.esigem.org>) que incluye 41 hospitales españoles. Los casos en este

estudio son 307 pacientes pediátricos con EM confirmada, lo que representa la mayor serie de pacientes de EM analizada hasta la fecha. Para la comparación, se emplearon dos colecciones independientes de muestras de controles (CG1 [N= 917] y CG2 [N= 616]), emparejados por etnicidad. Se analizó en casos y controles una selección de 25 SNPs del ADNmt (mtSNPs), algunos de ellos característicos de las principales ramas europeas de la filogenia mitocondrial. Además, se le genotiparon 34 marcadores informativos de ancestralidad (AIMs) a los casos y al grupo control CG2, con el objetivo de monitorizar la estratificación poblacional que pudiese estar oculta. Como grupos de clasificación de nuestros casos y controles se emplearon muestras conocidas de población con ancestralidad africana, nativo americana y europea ( $N = 711$ ). Un total de 39 individuos (incluidos 14 gitanos) fueron eliminados de los análisis estadísticos principales debido a su ascendencia (reportada *a priori* como) no española o a los resultados de los AIMs indicando una pertenencia europea menor del 95%. Los análisis de asociación de los restantes 268 casos contra CG1 sugirieron una sobrerrepresentación del alelo A del SNP sinónimo G11719A (test Chi-cuadrado de Pearson;  $P$ -valor ajustado= 0'0188; OR [95% IC]= 1'63 [1'22-2'18]. Sin embargo, cuando se compararon los casos frente a CG2, no se pudo replicar la asociación positiva. No se ha observado ninguna asociación debido a diferencias en la distribución de haplogrupos entre casos y controles, ni entre el haplogrupo y alguna variante clínica.

**Conclusiones:** No encontramos evidencias de asociación entre los mtSNPs y la enfermedad meningocócica después de monitorizar cuidadosamente el efecto de confusión de la sub-estructura poblacional. En las poblaciones humanas, la variabilidad del ADNmt está particularmente estratificada debido a su pequeño tamaño poblacional efectivo en comparación con marcadores autosómicos, por lo que se debe tener especial cuidado en la interpretación de señales que aparentemente indicarían asociaciones positivas en estudios de asociación caso-control del ADNmt.



















## **Artículo 5: Major mtDNA European Haplogroups are not Associated with Schizophrenia**

**Ana Mosquera-Miguel**, Manuel Arrojo, Eduardo Paz, Ramón Ramos-Ríos, Santiago Agra, Mario Páramo, Julio Brenlla, Silvia Martínez, Javier Costas, Antonio Salas

*Am J Med Genet Part B (Neuropsychiatric Genetics) Research Article  
(In preparation)*

### **Los principales haplogrupos mitocondriales europeos no están asociados con la esquizofrenia**

#### **ABSTRACT/RESUMEN:**

Estudios recientes han sugerido el papel de variantes mitocondriales comunes (mtSNPs) como factores de riesgo para la esquizofrenia. Sin embargo, todos estos estudios se han llevado a cabo con cohortes de pacientes relativamente pequeñas y la mayoría de ellos no han sido todavía replicados. El objetivo del presente trabajo es replicar estos resultados, por lo que se ha realizado el mayor esfuerzo de genotipado hasta la fecha empleando una cohorte de 454 pacientes españoles de esquizofrenia. Como grupo control se ha empleado una muestra de 616 individuos españoles no emparentados. No se ha observado ninguna asociación positiva entre mtSNPs o haplogrupos y la esquizofrenia. Además, también se discuten algunas de las deficiencias detectadas en trabajos previos que podrían explicar los falsos positivos de sus resultados, incluyendo el efecto de confusión debido a estratificación poblacional y las deficientes metodologías estadísticas. Es poco probable que variantes comunes del ADNmt estén relacionadas con la esquizofrenia, al menos en poblaciones europeas occidentales.

American Journal of Medical Genetics Part B (Neuropsychiatric Genetics)

*Research Article (In preparation)*

## **Major mtDNA European Haplogroups are not Associated with Schizophrenia**

**Ana Mosquera-Miguel<sup>1</sup>, Manuel Arrojo<sup>3</sup>, Eduardo Paz<sup>3</sup>, Ramón Ramos-Ríos<sup>3</sup>, Santiago Agra<sup>3</sup>, Mario Páramo<sup>3</sup>, Julio Brenlla<sup>3</sup>, Silvia Martínez<sup>3</sup>, Javier Costas<sup>2</sup>, Antonio Salas<sup>1\*</sup>**

<sup>1</sup> *Unidade de Xenética, Departamento de Anatomía Patolóxica e Ciencias Forenses and Instituto de Medicina Legal, Facultade de Medicina, Universidade de Santiago de Compostela, Galicia, Spain,*

<sup>2</sup> *Fundación Pública Galega de Medicina Xenómica (FPGMX)-SERGAS, Hospital Clínico Universitario, Universidade de Santiago; Galicia, Spain*

<sup>3</sup> *Servicio de Psiquiatría, Complejo Hospitalario Universitario de Santiago, Santiago de Compostela, Spain.*

**\*Correspondence to:** Antonio Salas, Unidade de Xenética, Departamento de Anatomía Patolóxica e Ciencias Forenses and Instituto de Medicina Legal, Facultade de Medicina, C/ San Francisco s/n; CP: 15706; Universidade de Santiago de Compostela, Galicia, Spain; Phone: +34 981 582 327; Fax: +00 981 580 336; E-mail: antonio.salas@usc.es

**Running title:** mtDNA haplogroups are not associated to SZ.

**Keywords:** mitochondrial DNA, SNPs, association studies, population stratification, false positive.

**Grant sponsors:** The present project was supported by grants from Xunta de Galicia (PGIDIT06PXIB208079PR and Grupos Emerxentes: 2008/037), Fundación de Investigación Médica Mutua Madrileña (2008/CL444) and Ministerio de Ciencia e Innovación (SAF2008-02971) given to AS.



## ABSTRACT

Recent studies have claimed a role of mtDNA common variants (mtSNPs) and/or haplogroups (hgs) to an increase risk to schizophrenia (SZ). All these studies however have mainly been undertaken on relatively small cohorts of patients and most have not yet been replicated. With the main aim of replicating these findings, we have carried out the largest genotyping effort to date using a cohort of 454 SZ Spanish patients. A sample consisting of 616 healthy Spanish unrelated individuals were used as control group. We did not observe any positive association between mtSNPs and hgs with SZ. We discussed several deficiencies of previous studies that might explain the false positive nature of previous findings, including the confounding effect of population sub-structure and deficient statistical methodologies. It is unlikely that common mtDNA variants are related to SZ, at least in Western European populations.

## INTRODUCTION

Schizophrenia (SZ) is a chronic, severe, and disabling brain disorder characterized by abnormalities in the perception or expression of reality and affecting about 0.4-0.6% of the population (Bhugra 2005). Symptoms rarely appear in childhood and include auditory hallucinations, paranoid delusions, disordered thinking, flat affect, cognitive deficits, movement disorders, social withdrawal, etc. There is considerable heterogeneity in clinical presentation across patients and the etio-pathogenesis of the disease is largely unknown. Genetics is one of the contributing factors to SZ together or through interaction with neurobiology and psychological processes or early environments. For instance, twin studies and adoption studies have suggested a high level of heritability (O'Donovan et al. 2003) of SZ. Recent genome-wide association studies on SZ discard the existence of common variants of moderate effect, and suggest that thousand of common alleles of very small effect may be involved in the genetic risk to SZ (Purcell et al. 2009; Stefansson et al. 2009; Shi et al. 2009).

As reviewed in Rollins et al. (2009), there are several suggestive coincidences that make the mtDNA an attractive target for the study of the genetic causes underlining SZ. Briefly, the mtDNA is inherited in a matrilineal fashion (Giles et al. 1980) while high rates of maternal offspring from schizophrenic patients compared to paternal offspring have been observed in family studies (Goldstein et al. 1992; Wolyniec et al. 1992). MtDNA mutations appear more frequently in tissues with higher metabolic rates, such as the brain and there are several studies suggesting small changes in the level of heteroplasmy in neurodegenerative disorders such as Parkinson's or Alzheimer's disease. Some studies have proposed several mtDNA common variants as risk factors in SZ (Amar et al. 2007; Rollins et al. 2009; Ueno et al. 2009). Magri et al. (2007) have recently suggested an association between mtDNA hg JT and the onset of SZ. All these studies however have mainly been undertaken on relatively small cohorts of patients and most have not been replicated yet. Recently, Doi et al. (2009) provided an evolutionary view that would support the role of mtDNA in SZ basically based on its high mutation rate compared to the lower average mutation rate in the nuclear DNA.

The present study represents the largest effort to date in terms of sample size aimed to unravel the presumable association of common mtDNA variants representing main West European mtDNA hgs to SZ.

## **MATERIAL AND METHODS**

### **Study subjects**

Cases were 454 schizophrenic patients attending the Galician Mental Health Service in the area of Santiago de Compostela (Galicia, NW Spain), meeting DSM-IV criteria for schizophrenia. All patients were Spaniards of European ancestry. Patients with mental retardation or comorbid neurological diseases were excluded. The control group ( $N = 616$ ) consisted of ethnicity-matched Spanish individuals and were previously reported as CG2 in Salas et al. (2009). Knowing that the prevalence of having SZ is about 0.5%, we would expect less than 31 false negatives among our controls.

Written informed consent was gathered from all patients and the samples were collected anonymously. The study and the use of archive samples for this project were approved by the Clinical Research Ethics Committee of Galicia and the Ethics Committee of the University of Santiago de Compostela. The study was conducted according to the Spanish Law for Biomedical Research (Law 14/2007- 3 of July) and complied with the Declaration of Helsinki.

### **mtSNP selection and genotyping**

Selection of mtDNA SNPs was carried out as in Mosquera-Miguel et al. (2008) and Salas et al. (2009). A set of 25 mtSNPs representing main branches of the European mtDNA phylogeny, were genotyped in cases and controls.

All the samples were genotyped according to the protocols described previously (Quintáns et al. 2004; Álvarez-Iglesias et al. 2008). The genotyping method was based on minisequencing SNaPshot technique (Álvarez-Iglesias et al. 2007). MtDNA haplotypes were checked by a phylogenetic-based approach as performed in previous studies (Bandelt et al. 2002; Bandelt et al. 2004a; Bandelt et

al. 2004b; Salas et al. 2005a; Salas et al. 2005b), in order to minimize the probability of genotyping errors.

### **Association tests**

Associations were assessed for individual mtSNPs by comparison of allele frequencies between cases and controls. We have used a one degree of freedom chi-square test, or Fisher's exact test for cell counts below five. We have used a nominal value of  $\alpha = 0.05$  to assess the significance of the association. Multiple testing was addressed using a permutation test.

Statistical analyses were carried out using the statistical packages Stata v.8 (<http://www.stata.com/>) and R (<http://www.r-project.org/>).

Association was also assessed for hgs by comparing the frequency of each hg *versus* the other hgs aggregated between cases and controls. A permutation test was also carried out to address the issue of multiple test correction.

According to Magri et al. (2007) there is a statistically significant correlation between hg JT carriers and the age at onset of schizophrenia. Following the same methodology employed by these authors, we used the Kaplan-Meier method and the log-rank test for the analysis of survival; both using the software PASW Statistics v. 18. Quanto software (Gauderman 2002) was used for power calculations. Note however that power was computed under allelic principles that do not necessarily fulfill the specific conditions affecting mtDNA variants (see also (Salas et al. 2009)).

## **RESULTS**

The power to detect odds ratios as low as 1.5 for mtSNPs and considering an average population minimum allele frequency (MAF) of 10% (lower than the average MAF in controls) was 84%. Pearson's chi-square or Fisher's exact tests were used to assess association. Adjustment of *P*-values for multiple tests was carried out by way of a permutation test (20,000 permutations) as previously implemented in Mosquera-Miguel et al. (2008) and Salas et al. (2009). When analyzing mtSNP association between cases and controls, we observed the best *P*-value (Pearson's  $\chi^2$  test, nominal *P*-value = 0.004) for mtSNP T6776C which defines

hg H3 (OR = 2.09; 95% CI = 1.26-3.47); a value which is statistically significant under a nominal value of  $\alpha = 0.05$ . Nevertheless, this apparent association is not sustained after correcting for multiple hypothesis using a permutation test (adjusted  $P$ -value = 0.0609) (Table I).

We did not find association of hgs *versus* SZ patients. In agreement with the association found for mtSNP T6776C; the best  $P$ -value was observed for hg H3 (Pearson's  $\chi^2$  test, nominal  $P$ -value = 0.010) with an OR of 2.09 (1.26-3.47; 95% CI). This  $P$ -value does not remain significant after correcting for multiple test comparisons (permutation test, adjusted  $P$ -value = 0.0946; Table II).

Magri et al. (2007) reported that schizophrenic patients harboring hg JT presented statistically significant (Log Rank  $\chi^2 = 10.298$ ,  $P$ -value = 0.016) earlier onset of schizophrenia as compared to patients carrying different hgs. Encouraged by this finding, we carried out the same analysis with our patients. As shown in Figure 1, there is not an apparent relationship between the age at onset and hg JT; and the Log Rank test did not suggest a statistical association when comparing hg JT carriers with the rest of the patients (Log Rank  $\chi^2 = 0.582$ ;  $P$ -value=0.446) or when considering the four hg categories used by Magri et al., that is, hg UK, TJ, H, and others (Log Rank  $\chi^2 = 0.698$ ;  $P$ -value=0.874). Magri et al. sub-divided the patients in two groups according to their age at onset (<28 years vs onset >28 years) and carried out a binary logistic regression analysis using as fixed factor mtDNA hg JT *versus* the remaining hgs. These authors found that hg JT had a 7.8 (95% CI 1.7-34.7) fold increased risk of developing schizophrenia before 28 years. As indicated by a logistic regression test, hg JT carriers do not have an increased risk of developing schizophrenia for a cut-off of 28 years; Wald's test = 1.600 (un-adjusted  $P$ -value = 0.206); and OR = 1.404 (95% CI: 0.83-2.38).

## DISCUSSION

It has been suggested that many genes might be contributing to SZ, couple with a complex interplay with environmental factors. Some recent studies have suggested several mtDNA hgs as predisposition factors in SZ. With the aim of replicating other's findings, we have carried out another study in a cohort of

Spanish SZ patients. The main West European mtDNA hgs were targeted in the present study by way of analyzing 25 different mtSNPs. We did not observed positive association of these mtSNPs or hgs with the SZ condition, which contradicts the findings observed by others. There are however several reasons that could explain the false positive nature of previous findings.

Amar et al. (2007) have recently investigated the role of common mtDNA hgs in SZ Israeli Arab patients ( $N=202$ ). A logistic regression test was carried out considering the disease status and hgs in cases and controls. Results were regarded as statistically significant if  $\alpha < 0.05$ . According to the authors, hg HV was overrepresented in cases with respect to control with a significant  $P$ -value of 0.01. There are at least two major drawbacks related to this study. First, the authors stated that *"the control group in the present study was comprised of the patients' psychiatrically healthy fathers, thus, it is unlikely that population stratification plays a confounding factor in this study"*; this is however a fallacy because there is not a full guaranty that the population genetic background of fathers fully represent the population background of mothers. The authors were perhaps confused by other tests traditionally employed in family-based association studies (e.g. TDT) but however used under the diploid properties of autosomal markers, which obviously do not apply to the mtDNA variation. More importantly, the authors did not adjust their  $P$ -values for multiple test comparisons, despite the fact of carrying out several association tests. For instance, similar unadjusted  $P$ -values obtained in other case-control mtDNA association studies yielded adjusted  $P$ -values very above (Mosquera-Miguel et al. 2008) to the significant level used by Amar et al. (2007) (see also Table I). Similar concerns regarding statistical analysis apply to the study of Rollins et al. (2009); for instance, these authors claimed that *"another exploratory difference in the distribution of psychiatric groups by haplogroup (Table 1) was seen in analysis of the combined BD+SZ distribution of subjects found more frequently in the pre HV group (Table 2) compared to all haplogroups (Chi square=13.7; df(6),  $p=0.032$ )"*; just by chance, it is expected to observed some seeming positive findings of association among the large number of hypothesis tested in their study. Moreover, stratification cannot be disregarded in the study by Rollins et al. (2009).

Magri et al. (2007) analyzed 142 SZ patients and 190 controls without any replication confirmatory sample. These authors claimed that “*patients harboring the hg J-T showed an anticipated onset of the disorder*” and that “*these results indicate that the J-T hg of mt-DNA may have a modulator effect on deeper determinants of schizophrenia*”. First, given the small sample size employed in this study, the power to detect any association is sub-optimum, which is more aggravated when the samples are stratified by age (see their Table II; there are only 9 and 15 females and males, respectively, belonging to JT). Second, no replication and confirmatory sample was employed to corroborate their findings and no control for population stratification was established. Our data, involving more than three times the sample size used in the Magri’s et al. study, does not indicate association between JT carriers and the age at onset of schizophrenia.

MtDNA variants are strongly stratified in human populations, as already observed in population genetic studies (Torroni et al. 2001; Loogväli et al. 2004; Achilli et al. 2004; Álvarez-Iglesias et al. 2009) and suggested in case-control disease studies (Salas et al. 2009; Mosquera-Miguel et al. 2008). Control for stratification is a great challenge because the level of stratification captured by autosomal SNPs are probably inadequate in mtDNA studies (Salas et al. 2009). Stratification acts in the direction of creating false positive findings, and therefore, can not affect the results and conclusions of the present study; but it could be contributing factor in other studies where positive findings were reported. As done in the analysis of autosomal predisposition factors, a replication study using an independent sample should be mandatory in order to prevent false claims of positive associations in the study of complex diseases.

Note that many journals, including the Part B of the *American Journal of Medical Genetics* have now specific editorial policies on association studies, including “*any positive finding would include an adequately powered replication samples*” and “*P-values should be corrected for multiple comparisons*” (<http://www3.interscience.wiley.com/journal/117928900/grouphome/ForAuthors.html>). According to these criteria and taking into account the results presented in this study, all the above mentioned positive finding should stand on ‘quarantine’.

In addition, several authors have proposed that rare variants in the mtDNA may convey susceptibility to SZ. Ueno et al. (2009) analyzed the complete genomes

of 93 Japanese schizophrenic patients. They observed three non-synonymous homoplasmic variants in subunit six of the MT-ATP6 gene of patients but not in controls that, according to the authors, were suggested to be slightly deleterious for two main reasons: their original amino acid residues were highly conserved, and the physicochemical differences between the original and altered amino acid residues were relatively high. In addition, these authors detected three novel heteroplasmic variants that were potentially pathogenic. Martorell et al. (2006) have also contributed to the analysis of rare variants by way of analyzing six schizophrenic patients with apparent maternal transmission of SZ. The potential pitfalls related to the analysis of rare mtDNA variants in SZ were already discussed (Bandelt et al. 2007; Bandelt et al. 2005) and deserves further attention in future studies.

In conclusion, our results, supported on the largest cohort of schizophrenic patients analyzed to date, suggest that major European mtDNA hgs are not risk factors in SZ. There are reasons to believe that previous findings do not rest on a solid ground due to the presumable confounding effect of population stratification and the lack of proper statistical analyses.

## ACKNOWLEDGEMENTS

We thank the patients for their participation, and the healthcare professionals of Complejo Hospitalario de Santiago who assisted in the recruitment of study participants.

## REFERENCES

- Achilli, A., C. Rengo, C. Magri, V. Battaglia, A. Olivieri, R. Scozzari, F. Cruciani, M. Zeviani, E. Briem, V. Carelli, P. Moral, J. M. Dugoujon, U. Roostalu, E. L. Loogvali, T. Kivisild, H. J. Bandelt, M. Richards, R. Villems, A. S. Santachiara-Benerecetti, O. Semino, and A. Torroni. 2004. The molecular dissection of mtDNA haplogroup H confirms that the Franco-Cantabrian glacial refuge was a major source for the European gene pool. *Am J Hum Genet* 75 (5):910-918.
- Álvarez-Iglesias, V., F. Barros, Á. Carracedo, and A. Salas. 2008. Minisequencing mitochondrial DNA pathogenic mutations. *BMC Med Genet* 9:26.



- Álvarez-Iglesias, V., J. C. Jaime, Á. Carracedo, and A. Salas. 2007. Coding region mitochondrial DNA SNPs: targeting East Asian and Native American haplogroups. *Forensic Sci Int Genet* 1:44-55.
- Álvarez-Iglesias, V., A. Mosquera-Miguel, M. Cerezo, B. Quintáns, M. T. Zarrabeitia, I. Cuscó, M. V. Lareu, O. García, L. Pérez-Jurado, Á. Carracedo, and A. Salas. 2009. New population and phylogenetic features of the internal variation within mitochondrial DNA macro-haplogroup R0. *PLoS ONE* 4 (4):e5112.
- Amar, S., A. Shamir, O. Ovadia, M. Blanaru, A. Reshef, I. Kremer, M. Rietschel, T. G. Schulze, W. Maier, R. H. Belmaker, R. P. Ebstein, G. Agam, and D. Mishmar. 2007. Mitochondrial DNA HV lineage increases the susceptibility to schizophrenia among Israeli Arabs. *Schizophr Res* 94 (1-3):354-358.
- Andrews, R. M., I. Kubacka, P. F. Chinnery, R. N. Lightowlers, D. M. Turnbull, and N. Howell. 1999. Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. *Nat Genet* 23:147.
- Bandelt, H.-J., A. Olivieri, C. M. Bravi, Y.-G. Yao, A. Torroni, and A. Salas. 2007. 'Distorted' mitochondrial DNA sequences in schizophrenic patients. *Eur J Hum Genet* 15:400-404.
- Bandelt, H.-J., Y. G. Yao, and T. Kivisild. 2005. Mitochondrial genes and schizophrenia. *Schizophr Res* 72 (2-3):267-269.
- Bandelt, H. J., L. Quintana-Murci, A. Salas, and V. Macaulay. 2002. The fingerprint of phantom mutations in mitochondrial DNA data. *Am J Hum Genet* 71 (5):1150-1160.
- Bandelt, H. J., A. Salas, and C. Bravi. 2004a. Problems in FBI mtDNA database. *Science* 305 (5689):1402-1404.
- Bandelt, H. J., A. Salas, and S. Lutz-Bonengel. 2004b. Artificial recombination in forensic mtDNA population databases. *Int J Legal Med* 118 (5):267-273.
- Bhugra, D. 2005. The global prevalence of schizophrenia. *PLoS Med* 2 (5):e151; quiz e175.
- Doi, N., Y. Hoshi, M. Itokawa, C. Usui, T. Yoshikawa, and H. Tachikawa. 2009. Persistence criteria for susceptibility genes for schizophrenia: a discussion from an evolutionary viewpoint. *PLoS ONE* 4 (11):e7799.
- Gauderman, W. J. 2002. Sample size requirements for matched case-control studies of gene-environment interaction. *Stat Med* 21 (1):35-50.

- Giles, R. E., H. Blanc, H. M. Cann, and D. C. Wallace. 1980. Maternal inheritance of human mitochondrial DNA. *Proc Natl Acad Sci U S A* 77 (11):6715-6719.
- Goldstein, J. M., S. V. Faraone, W. J. Chen, and M. T. Tsuang. 1992. Gender and the familial risk for schizophrenia. Disentangling confounding factors. *Schizophr Res* 7 (2):135-140.
- Loogväli, E.-L., U. Roostalu, B. A. Malyarchuk, M. V. Derenko, T. Kivisild, E. Metspalu, K. Tambets, M. Reidla, H.-V. Tolk, J. Parik, E. Pennarun, S. Laos, A. Lunkina, M. Golubenko, L. Barač, M. Peričić, O. P. Balanovsky, V. Gusar, E. K. Khusnutdinova, V. Stepanov, V. Puzyrev, P. Rudan, E. V. Balanovska, E. Grechanina, C. Richard, J. P. Moisan, A. Chaventré, N. P. Anagnou, K. I. Pappa, E. N. Michalodimitrakis, M. Claustres, M. Gölge, I. Mikerezi, E. Usanga, and R. Villems. 2004. Disuniting uniformity: a pied cladistic canvas of mtDNA haplogroup H in Eurasia. *Mol Biol Evol* 21 (11):2012-2021.
- Magri, C., R. Gardella, S. D. Barlati, P. Valsecchi, E. Sacchetti, and S. Barlati. 2007. Mitochondrial DNA haplogroups and age at onset of schizophrenia. *Am J Med Genet B Neuropsychiatr Genet* 144B (4):496-501.
- Martorell, L., T. Segues, G. Folch, J. Valero, J. Joven, A. Labad, and E. Vilella. 2006. New variants in the mitochondrial genomes of schizophrenic patients. *Eur J Hum Genet* 14 (5):520-528.
- Mosquera-Miguel, A., V. Álvarez-Iglesias, A. Vega, R. Milne, A. Cabrera de León, J. Benitez, Á. Carracedo, and A. Salas. 2008. Is mitochondrial DNA variation associated with sporadic breast cancer risk? *Cancer Res* 68 (2):623-625.
- O'Donovan, M. C., N. M. Williams, and M. J. Owen. 2003. Recent advances in the genetics of schizophrenia. *Hum Mol Genet* 12 Spec No 2:R125-133.
- Purcell, S. M., N. R. Wray, J. L. Stone, P. M. Visscher, M. C. O'Donovan, P. F. Sullivan, and P. Sklar. 2009. Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature* 460 (7256):748-752.
- Quintáns, B., V. Álvarez-Iglesias, A. Salas, C. Phillips, M. V. Lareu, and Á. Carracedo. 2004. Typing of mitochondrial DNA coding region SNPs of forensic and anthropological interest using SNaPshot minisequencing. *Forensic Sci Int* 140 (2-3):251-257.
- Rollins, B., M. V. Martin, P. A. Sequeira, E. A. Moon, L. Z. Morgan, S. J. Watson, A. Schatzberg, H. Akil, R. M. Myers, E. G. Jones, D. C. Wallace, W. E. Bunney, and

- M. P. Vawter. 2009. Mitochondrial variants in schizophrenia, bipolar disorder, and major depressive disorder. *PLoS One* 4 (3):e4913.
- Salas, A., A. Carracedo, V. Macaulay, M. Richards, and H. J. Bandelt. 2005a. A practical guide to mitochondrial DNA error prevention in clinical, forensic, and population genetics. *Biochem Biophys Res Commun* 335 (3):891-899.
- Salas, A., L. Fachal, S. Marcos-Alonso, A. Vega, F. Martín-Torres, and G. ESIGEM. 2009. Investigating the role of mitochondrial haplogroups in genetic predisposition to meningococcal disease. *PLoS ONE* 4 (12):e8347.
- Salas, A., Y. G. Yao, V. Macaulay, A. Vega, A. Carracedo, and H. J. Bandelt. 2005b. A critical reassessment of the role of mitochondria in tumorigenesis. *PLoS Med* 2 (11):e296.
- Shi, J., D. F. Levinson, J. Duan, A. R. Sanders, Y. Zheng, I. Pe'er, F. Dudbridge, P. A. Holmans, A. S. Whittemore, B. J. Mowry, A. Olincy, F. Amin, C. R. Cloninger, J. M. Silverman, N. G. Buccola, W. F. Byerley, D. W. Black, R. R. Crowe, J. R. Oksenberg, D. B. Mirel, K. S. Kendler, R. Freedman, and P. V. Gejman. 2009. Common variants on chromosome 6p22.1 are associated with schizophrenia. *Nature* 460 (7256):753-757.
- Stefansson, H., R. A. Ophoff, S. Steinberg, O. A. Andreassen, S. Cichon, D. Rujescu, T. Werge, O. P. Pietilainen, O. Mors, P. B. Mortensen, E. Sigurdsson, O. Gustafsson, M. Nyegaard, A. Tuulio-Henriksson, A. Ingason, T. Hansen, J. Suvisaari, J. Lonnqvist, T. Paunio, A. D. Borglum, A. Hartmann, A. Fink-Jensen, M. Nordentoft, D. Hougaard, B. Norgaard-Pedersen, Y. Bottcher, J. Olesen, R. Breuer, H. J. Moller, I. Giegling, H. B. Rasmussen, S. Timm, M. Mattheisen, I. Bitter, J. M. Rethelyi, B. B. Magnusdottir, T. Sigmundsson, P. Olason, G. Masson, J. R. Gulcher, M. Haraldsson, R. Fossdal, T. E. Thorgeirsson, U. Thorsteinsdottir, M. Ruggeri, S. Tosato, B. Franke, E. Strengman, L. A. Kiemeny, I. Melle, S. Djurovic, L. Abramova, V. Kaleda, J. Sanjuan, R. de Frutos, E. Bramon, E. Vassos, G. Fraser, U. Ettinger, M. Picchioni, N. Walker, T. Touloupoulou, A. C. Need, D. Ge, J. L. Yoon, K. V. Shianna, N. B. Freimer, R. M. Cantor, R. Murray, A. Kong, V. Golimbet, A. Carracedo, C. Arango, J. Costas, E. G. Jonsson, L. Terenius, I. Agartz, H. Petursson, M. M. Nothen, M. Rietschel, P. M. Matthews, P. Muglia, L. Peltonen, D. St Clair, D. B. Goldstein, K. Stefansson,

- and D. A. Collier. 2009. Common variants conferring risk of schizophrenia. *Nature* 460 (7256):744-747.
- Torrioni, A., H.-J. Bandelt, V. Macaulay, M. Richards, F. Cruciani, C. Rengo, V. Martínez-Cabrera, R. Villems, T. Kivisild, E. Metspalu, J. Parik, H. V. Tolk, K. Tambets, P. Forster, B. Karger, P. Francalacci, P. Rudan, B. Janicijevic, O. Rickards, M. L. Savontaus, K. Huoponen, V. Laitinen, S. Koivumaki, B. Sykes, E. Hickey, A. Novelletto, P. Moral, D. Sellitto, A. Coppa, N. Al-Zaheri, A. S. Santachiara-Benerecetti, O. Semino, and R. Scozzari. 2001. A signal, from human mtDNA, of postglacial recolonization in Europe. *Am J Hum Genet* 69 (4):844-852.
- Ueno, H., Y. Nishigaki, Q. P. Kong, N. Fuku, S. Kojima, N. Iwata, N. Ozaki, and M. Tanaka. 2009. Analysis of mitochondrial DNA variants in Japanese patients with schizophrenia. *Mitochondrion* 9 (6):385-393.
- Wolyniec, P. S., A. E. Pulver, J. A. McGrath, and D. Tam. 1992. Schizophrenia: gender and familial risk. *J Psychiatr Res* 26 (1):17-27.

**Table I.** Pearson's chi-square test for cases *versus* controls.

SNP	rCRS <sup>a</sup>	MAF controls <sup>c</sup>	MA <sup>b</sup>	MAF cases <sup>c</sup>	Chi2 exact	P-value	Adjusted P-value <sup>d</sup>	OR (95% CI) <sup>e</sup>
G3010A	G	0.72	A	0.24	1.9292	0.165	0.9497	0.82 (0.62-1.09)
T3197C	T	0.10	C	0.09	0.8584	0.354	0.9996	1.22 (0.80-1.87)
G3915A	G	0.96	A	0.02	1.7797	0.182	0.9681	0.61 (0.30-1.27)
C3992T	C	0.01	T	0.01	0.1665	0.683	1	0.77 (0.23-2.66)
T4216C	T	0.16	C	0.16	0.0000	0.999	1	1.00 (0.72-1.40)
T4336C	T	0.04	C	0.04	0.0961	0.757	1	0.90 (0.47-1.72)
A4529T	A	0.00	T	0.01	2.8892	0.089	0.7895	5.45 (0.61-48.92)
G4580A	G	0.03	A	0.05	2.7831	0.095	0.7971	1.67 (0.91-3.06)
A4769G	A	0.03	A	0.02	0.9891	0.610	1	1.21 (0.57-2.60)
A4793G	A	0.01	G	0.00	2.1452	0.143	0.9241	0.23 (0.03-1.95)
T6776C	T	0.04	C	0.09	8.4279	0.004	0.0609	2.09 (1.26-3.47)
C7028T	C	0.45	T	0.44	0.1156	0.734	1	1.04 (0.82-1.33)
G8994A	G	0.02	A	0.02	0.5273	0.468	1	1.39 (0.57-3.36)
G9055A	G	0.10	A	0.10	0.0004	0.983	1	1.00 (0.66-1.52)
A10398G	A	0.20	G	0.19	0.0238	0.877	1	0.98 (0.72-1.33)
C10400T	C	0.01	T	0.00	1.0068	0.316	0.9983	0.45 (0.09-2.24)
T10463C	T	0.09	C	0.10	0.1176	0.732	1	1.08 (0.70-1.66)
C10873T	C	0.04	C	0.03	0.1320	0.716	1	1.13 (0.58-2.19)
G11719A	G	0.47	A	0.47	0.0676	0.795	1	0.97 (0.76-1.24)
A12308G	A	0.24	G	0.22	0.5201	0.471	1	0.99 (0.67-1.20)
C12705T	C	0.07	T	0.09	1.0049	0.316	0.9983	1.26 (0.80-1.96)
G13708A	G	0.09	A	0.10	0.3373	0.561	1	1.13 (0.75-1.70)
A13966G	A	0.02	G	0.03	0.3838	0.536	1	1.27 (0.59-2.74)
C14766T	C	0.48	C	0.47	0.1121	0.738	1	0.96 (0.75-1.22)
T16519C	T	0.35	T	0.33	0.2561	0.613	1	1.07 (0.83-1.38)

<sup>a</sup>rCRS: allele in the revised Cambridge Reference Sequence (rCRS) (Andrews et al. 1999);

<sup>b</sup>MA: allele with the lowest frequency;

<sup>c</sup>MAF: minimum allele frequency computed on control individuals;

<sup>d</sup>Adjusted P-value: adjustment of chi-square P-values was carried out with a permutation-based approach; number of permutations = 20,000;

<sup>e</sup>OR (95%CI): ORs were computed with the rCRS allele as a reference.

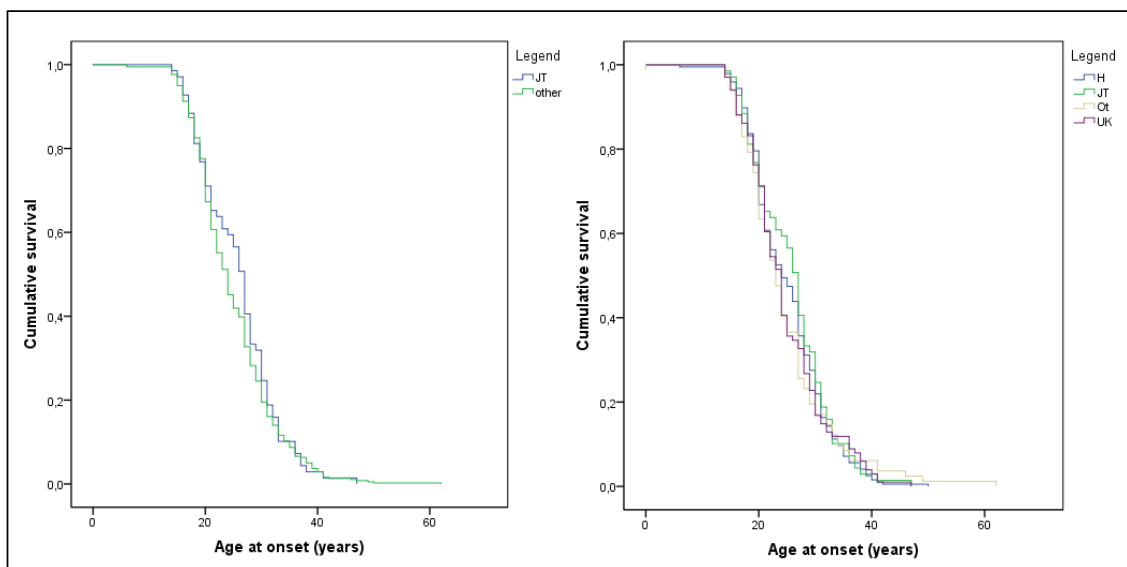
**Table II.** Pearson's chi-square test for the hg status of Galician cases *versus* controls. The test was carried out for those (sub)hgs with frequencies above 5% in the control groups as a first criteria but also for the well-known branches of the West European phylogeny I, V, and W.

HG	Freq. Cases	Freq. Controls	Chi2 exact	P-value	adjusted P-value <sup>a</sup>	OR (95% CI) <sup>b</sup>
H	0.44	0.45	0.0695	0.792	1	1.03 (0.81-1.32)
H1	0.19	0.21	0.8530	0.356	0.987	1.16 (0.85-1.57)
H3	0.09	0.05	6.6376	0.010	0.095	2.09 (1.26-3.47)
HV	0.52	0.48	0.0576	0.810	1	0.97 (0.76-1.24)
I	0.01	0.00	2.9028	0.088	0.618	0.18 (0.02-1.64)
J	0.07	0.07	0.0002	0.989	1	1.00 (0.63-1.58)
J1	0.07	0.06	0.0002	0.989	1	1.00 (0.63-1.58)
K	0.08	0.09	0.3829	0.536	0.999	1.15 (0.74-1.79)
K1	0.07	0.08	0.2090	0.648	1	1.11 (0.70-1.77)
R	0.91	0.93	1.0530	0.305	0.973	1.26 (0.81-1.96)
R0	0.53	0.51	0.0441	0.834	1	0.97 (0.76-1.24)
T	0.08	0.08	0.0058	0.939	1	1.02 (0.65-1.58)
TJ	0.16	0.16	0.0023	0.962	1	1.01 (0.72-1.41)
U	0.22	0.24	0.3838	0.536	0.999	1.10 (0.82-1.46)
U5	0.08	0.10	1.8670	0.172	0.859	1.35 (0.88-2.06)
V	0.05	0.03	2.7574	0.090	0.632	0.60 (0.33-1.10)
W	0.01	0.01	0.0029	0.957	1	1.03 (0.36-3.27)

<sup>a</sup>Adjusted P-value: adjustment of chi-square P-values was carried out with a permutation-based approach; number of permutations = 20,000;

<sup>b</sup>OR (95%CI): ORs were computed using as a reference the corresponding haplogroup frequencies independently to their frequencies.

**Figure 1.** Kaplan-Meier plots showing the age at onset in schizophrenic patients carrying the hg TJ *versus* the other patients (left) and following the same hg subdivision of Magri et al. (2007), namely, hgs H, TJ, UK and Other (right).



## **Artículo 6: Reassessing the role of mitochondrial DNA mutations in autism**

Álvarez-Iglesias V, **Mosquera-Miguel A**, Ivon Cuscó, Carracedo Á, Pérez-Jurado L, Salas A.

*American Journal of Medical Genetics Part B (Neuropsychiatric Genetics)*

*(In preparation)*

### **Reevaluación del papel de las mutaciones del ADN mitocondrial en el autismo**

#### **ABSTRACT/RESUMEN:**

Cada vez existen más evidencias sobre el importante papel de la disfunción del metabolismo energético mitocondrial en la etiología de los desórdenes del espectro autista (ASD, OMIM: 209850). Varios estudios han reivindicado que mutaciones en la molécula de ADN mitocondrial (ADNmt) podrían estar implicadas en el fenotipo de la enfermedad. Existen numerosas razones para creer que estas afirmaciones podrían ser todavía prematuras y por lo tanto la asociación entre mutaciones del ADNmt y los ASD están abiertas a debate. En este estudio se ha analizado una cohorte de 158 pacientes para una serie de mutaciones propuestas en la literatura como patogénicas en ASD. Además, se ha llevado a cabo un estudio de asociación caso-control para los haplogrupos europeos más frecuentes y para sus polimorfismos diagnóstico mediante la comparación de los casos con 753 controles sanos, emparejados por etnicidad. No se encontró ninguna evidencia de asociación entre mutaciones o polimorfismos del ADNmt en autismo, por lo que se concluye que los hallazgos previos de asociación entre variantes del ADNmt y ASD fueron falsos positivos. Los resultados de este trabajo son compatibles con la idea de que la frecuente observación de casos concomitantes de autismo y enfermedades mitocondriales pueda deberse a un factor nuclear que afecte a la función mitocondrial o a interacciones más complejas entre el núcleo y la mitocondria o el ADNmt.

American Journal of Medical Genetics Part B (Neuropsychiatric Genetics)

*(In preparation)*

## **Reassessing the role of mitochondrial DNA mutations in autism**

Álvarez-Iglesias V<sup>1</sup>, Mosquera-Miguel A<sup>1</sup>, Ivon Cuscó<sup>2</sup>, Carracedo Á<sup>1</sup>, Pérez-Jurado L<sup>2,3</sup>, Salas A<sup>1\*</sup>

**1** Unidade de Xenética, Instituto de Medicina Legal and Departamento de Anatomía Patolóxica e Ciencias Forenses, Facultade de Medicina, Universidade de Santiago de Compostela, Galicia, Spain

**2** Unidad de Genética, Universitat Pompeu Fabra, and U735 CIBER de enfermedades raras (CIBERER), Barcelona, Spain

**3** Programa de Medicina Molecular y Genética, Hospital Universitari Vall d'Hebron, Barcelona, Spain

**\*Correspondence to:** Antonio Salas, Unidade de Xenética, Departamento de Anatomía Patolóxica e Ciencias Forenses and Instituto de Medicina Legal, Facultade de Medicina, C/ San Francisco s/n; CP: 15706; Universidade de Santiago de Compostela, Galicia, Spain; Phone: +34 981 582 327; Fax: +00 981 580 336; E-mail: antonio.salas@usc.es

**Running title:** mtDNA variants in ASD.



**Abstract**

There is compelling evidence that impairment of mitochondrial energy metabolism play an important role in the aetiology of Autism Spectrum Disorders (ASD; OMIM number: 209850). Several studies claimed that mutations in the mitochondrial DNA (mtDNA) molecule could be involved in the disease phenotype. There are a number of reasons to believe that these claims could be still premature and therefore, the link between mtDNA mutations and ASD is open to debate. We have analyzed a cohort of 158 patients for a number of mutations proposed in the literature as pathogenic in ASD. We have also carried out a case control association study for the most common European haplogroups and their diagnostic polymorphisms by comparing cases with 753 healthy and ethnicity matched controls. We did not find evidence for an association between mtDNA mutations or polymorphisms in autism. We conclude that previous findings of association between mtDNA variants and ASD are spurious positives. Our results are compatible with the idea that the frequent observation of concomitant cases of autism and mtDNA disorders could be due to a nuclear factor influencing the mitochondrion function or due to a more complex interplay between the nucleus and the mitochondrion/mtDNA.

**Keywords:** autism spectrum disorders, haplogroup, mtDNA mutation

## Introduction

Autism/autistic Spectrum Disorders (ASD) are complex neuro-developmental conditions affecting one each 200 children. It is characterized by disturbance in language, perception and socialization. Autism has been associated with metabolic disorders in several recent studies (Oliveira et al. 2005; Correia et al. 2006), and there is increasing evidence that ASD have an important genetic component (Ramos et al. 2004). Linkage to a number of different loci has been reported as well as a number of candidate genes (Folstein and Rosen-Sheidley 2001; Ramos et al. 2004; Veenstra-VanderWeele and Cook 2004).

Recent studies have also suggested that impairment of mitochondrial energy metabolism play a role in the aetiology of autism (Chugani et al. 1999; Laszlo et al. 1994; Minshew et al. 1993). The study by Oliveira et al. (2005) found that 7% of a population-based survey of school-age children with ASD met criteria for mitochondrial respiratory chain disorders although they were clinically indistinguishable from other children with ASD. This already recognized feature of ASD has lead several researchers to analyze mutations at the mtDNA as potential risk factors in ASD. Thus, Graf et al. (2000) reported a family with a heterogeneous group of neurologic disorders associated with the mtDNA 8363G>A transfer ribonucleic acid (RNA)<sup>Lys</sup> mutation; the phenotype of one child in the family was consistent with autism. Pons et al. (2004) have explored the association of mtDNA mutations with autistic spectrum disorders, with special emphasis on mutation A3243G and mtDNA depletion. According to these authors, autistic spectrum disorders with or without additional neurologic features can be early presentation of the A3243G mutation and can be a prominent clinical manifestation of mtDNA depletion.

There is compiling evidence that mtDNA inherited sequence variants also have a subtle influence on respiratory chain activity during a critical stage of neurodevelopment which is highly energy dependent; therefore, these variants could contribute to the complex aetiology of neuro-developmental disorders such as ASD. With this hypothesis in mind, Kent et al. (2008) have investigated the most common mtDNA variants defining the main European haplogroups (clades of maternal lineages phylogenetically close related) in autism probands *versus* two

healthy control groups. Their results showed no evidence of an association of any mtDNA hg in autism. One could have in mind however the possibility of a type II error in the study of Kent et al. (2008) due to the limited statistical power for detecting an association provided by their sample size of 162 probands.

Using dHPLC technology, Weissman et al. (2008) have reported mtDNA mutations in 11 of 25 ASD patients. The mutations reported were 3397A>G, 4295A>G (both regarded as variants of probable pathogenicity), and 3394T>C, 10394C>T, 11809T>C, and 11984T>C (regarded as variants of unclear pathogenicity) (see their Table 4). Pathogenicity in this study was evaluated “*by search of the MITOMAP, and mtDB-Human Mitochondrial Genome database...*”. It is not clear however whether these mutations can be regarded as pathogenic. For instance, the scoring system proposed by Mitchell et al. (2006) gives to 3397A>G a low score, while its “pathogenic” status in MITOMAP is regarded as provisional. Phylogenetic considerations do not favor its disease status (Bandelt et al. 2006; Bandelt et al. 2009; Bandelt et al. 2008): this variant appears independently in several branches of the phylogeny in apparently healthy individuals (<http://www.phylotree.org/>).

There are therefore a number of evidences that seems to suggest the role of mtDNA variants in ASD; however, the evidences are still weak and have still not been replicated in a different cohort of patients. The main aim of the present work is to explore the potential association of mtDNA mutations/polymorphisms in both control and coding regions and/or haplogroup background in autism.

## **Material and Methods**

### *Patients*

The characteristics of a subset of the patients used in the present study are described in Cuscó et al. (2009). Briefly, a total of 158 Spanish patients (92 children followed in the neurology clinic and 66 institutionalized mentally retarded adults) with a confirmed diagnosis of one of the categories of ASD listed in the Diagnosis and Statistical Manual of Mental Diseases (DSM-IV) were analyzed. All patients were studied using the ADI-R to define a specific category of ASD and the Wechsler Intelligence Scale for Children or Wechsler Adult Intelligence Scale,

allowing us to measure general, verbal and performance IQ as well as analyses of multiple factorial components of cognitive functioning. The Leiter International Performance Scale-Revised and the Raven Progressive Matrices in the non-verbal patients were also used. All patients had an extensive evaluation by neurologists and clinical geneticists along with an intensive laboratory workup including standard karyotyping, fragile X molecular testing, subtelomeric and targeted MLPA assays (homemade panel designed to detect genomic duplications/deletions of specific regions associated with ASD and mental retardation: 1p36, 2q37, 7q11.23, 15q11-q13, 17p11.2, 16p11.2, 22q11.2 and 22q13.3), as well as metabolic and brain image studies in some cases when clinically indicated (Gener et al. submitted). Table I summarizes some medical and demographic features of the ASD patients collected in the present study. Mitochondrial clinical alterations were not detected for none of the patients.

#### *Control group*

Two different control groups were used in the present study. A control group of DNA samples from 137 population control individuals matched for population ancestry (Spanish anonymous blood donors) were collected and used to analyzed presumable pathogenic mutations. The second control group referred as CG2 used in Salas et al. (2009) consists of 616 healthy Spanish individuals. The two control groups were combined together ( $N=753$ ) for the case-control association study involving the genotyping of polymorphisms that define the main European haplogroups.

#### *DNA extraction*

Genomic DNA was extracted from blood samples using a salting out method and employing the Puregene DNA Purification Kit (Gentra Systems).

#### *Ethical approval*

All subjects participated after written informed consent was obtained from their families or other legal caregivers. DNA extracts were submitted to the laboratory in Santiago de Compostela where the genotyping was carried out. In addition, the study was approved by the Ethical committee of the University of

Santiago de Compostela. The study conforms to the Spanish Law for Biomedical Research (Law 14/2007- 3 of July).”

#### *Automatic sequencing*

The mtDNA hypervariable region I (HVS-I) segment was sequenced in forward and reverse directions according to the protocol described in Álvarez-Iglesias et al. (2009).

#### *Genotyping of mtDNA variants*

The following 25 well-known pathogenic mutations were genotyped in all the patients: 3243A>G; 3460G>A; 3697G>A, 3946G>A, 3949T>C, 7445A>G/C, 8993T>G/C, 9176T>C/G, 10158T>C, 10191T>C, 10663T>C, 11777C>A, 11778G>A, 11832G>A, 12706T>C, 13513G>A, 13514A>G, 14459G>A, 14482C>A/G, 14484T>C, and 14487T>C. We have used the minisequencing assay reported in Álvarez-Iglesias et al. (2008).

In addition, the six mtDNA mutations reported by Weissman et al. (2008) as probable or unclear pathogenicity, namely, 3397A>G, 4295A>G, 3394T>C, 10394C>T, 11809T>C, and 11984T>C, were also genotyped. A new minisequencing assay was designed *ad hoc*, following the methodology reported before (Álvarez-Iglesias et al. 2008; Mosquera-Miguel et al. 2009).

The hg status of several mtDNA profiles cannot be inferred solely from control region data. We therefore genotyped the mtSNPs defining the main branches of the European mtDNA phylogeny as done in Quintáns et al. (2004) and Mosquera-Miguel et al (2008). These polymorphisms were used for a case-control association study.

#### *Association study*

A case-control association study was carried out between ASD patients ( $N=158$ ) and controls ( $N=753$ ) for hg status and individually for each mtSNPs. Firstly, allele frequencies between cases and controls were compared in order to assess individual mtSNP association and using a one degree of freedom chi-square test (or Fisher’s exact test for cell counts below five). A nominal value of  $\alpha = 0.05$

was selected to assess the significance of the association. Secondly, the same statistical tests were used to carry out association analyses for hgs in cases and controls by comparing the frequency of each hg *versus* the other hgs aggregated. A permutation test was employed to address the issue of multiple testing.

Statistical analyses were carried out using the statistical packages Stata v.8 (<http://www.stata.com/>) and R (<http://www.r-project.org/>).

Quanto software (Gauderman 2002) was used for power calculations. See Salas et al. (2009) for some caveats.

## **Results**

### *Analysis of mtDNA mutations in ASD patients*

The mtDNA control region was sequenced in patients and a control group that consist of 137 healthy individuals (see M&M). Results are reported in Table S1. All these individuals and a second Spanish control group (N=616) were genotyped for a set of mtSNPs diagnostic of main mtDNA European hgs. All this information together indicates that the patients represent well the frequency pattern of a typical Spanish population (Álvarez-Iglesias et al. 2009; Salas et al. 1998). Thus, the most common haplogroup in cases and controls is by far hg H (~44%), where H1 is its most prevalent sub-lineage (~20%). Other well-known haplogroups are present, such as J, T, V, W, etc. We however observed in our patients some non-European lineages. Although there are not coding region SNPs to assist hg classification, some of these non-European mtDNAs can be allocated to typical East Asian lineages; however, given the fact that immigration from South America to Spanish cosmopolitan cities (such as Barcelona) has significantly increased during the last decade, it is highly probable that these Asian lineages belong to the Native American branches of the Asian phylogeny, that is, haplogroup A2, B2, C1, and D1. In addition, some individuals could be undoubtedly classified into typical sub-Saharan lineages belonging to different L-hgs (Table S1). Although we did not detect statistical association between cases and controls (see below), it is important to be aware of these outliers in cohorts of patients and controls because these non-European lineages could have been the source of a false positive finding of association (due to population stratification).

In addition, a total of 25 well-known pathogenic mtDNA mutations were minisequenced in all patients and 137 controls. All patients were carriers of non-pathogenic wild variants. We also genotyped the six suspected pathogenic mutations reported in Weissman et al. (2008); only 3397A>G was observed in one patient. Given the fact that the pathogenic status of these variants is debatable, it was not surprising to see 3394T>C in two healthy controls.

#### *Case-control association study*

A case-control association study was carried out for 158 patients and 753 controls. The mtSNPs genotyped represent well-known branches of the West European mtDNA phylogeny. Pearson's chi-square or Fisher's exact tests were used to assess mtSNP association. The best *P*-value (Pearson's  $\chi^2$  test, nominal *P*-value = 0.016) was found for mtSNP 12705C>T that lead from macro-hg N to R; this mtSNP apparently increases the risk of suffering ASD as indicated by the OR value (OR = 1.88; 95% CI = 1.12-3.18); however, the significance does not maintain after correcting for multiple hypothesis using a permutation test procedure (adjusted *P*-value = 0.1933) (Table II) as done in Mosquera-Miguel et al. (2008) and Salas et al. (2009).

Association test was carried out only for the most frequent haplogroups in order to increase the probability to detect any association given the limited sample size used in the present study (which is however high given the low prevalence of ASD; and when compared our sample size to other genetic articles on ASD). The frequency of mtSNPs and hgs in Iberia is already known from empirical population studies (Álvarez-Iglesias et al. 2009). The *a priori* power to detect odds ratios as low as 2 for mtSNPs and hgs with frequencies higher than 0.20% is about 96%, but the power decreases very significantly to 76% when frequencies are about 0.05%. A statistical power of 80% and OR = 1.5 can only be realistically obtained for the most frequent mtSNPs or hgs.

#### **Discussion**

Brain is strongly dependent on ATP production of the cell energy-producing organelle, the mitochondrion. Therefore, adequate mitochondrial metabolism is essential for normal brain function. There is a large body of evidence involving

mitochondrial dysfunction in autism; see Lombard (1998) for a concise review. While some evidences point to candidate nuclear DNA (nDNA) loci influencing mitochondrial function (Filipek et al. 2003; Oliveira et al. 2005; Ramoz et al. 2004), other seem to indicate the implications of mtDNA mutations as potential causative of autism (Pons et al. 2004). Although the implication of nDNA factors has found a relatively strong support, the presumable role of the mtDNA genome is still weak. The present study aims to provide more evidences supporting the role of mtDNA mutations or polymorphisms in ASD by screening one of the largest cohorts of patients analyzed to date. We therefore tried to screen the mtDNA of these patients for a wide spectrum of mutations and polymorphisms in order to explore the presumable role of high penetrance mtDNA mutations or low penetrance of common polymorphisms in autism.

We have targeted 25 well-known high penetrant pathogenic mutations in the mtDNA genome. All these mutations are common responsible mutations for a wide spectrum of mitochondrion diseases. We did not observed any of these mutations, either in homoplasmy or heteroplasmy condition, in the patients or controls analyzed in the present study. It is therefore unlikely that well known pathogenic mutations are responsible for autism disorders.

In addition, a set of other six presumable pathogenic mutations (Weissman et al. 2008) were also interrogated in our patients. We did not observe any of these mutations in our patients but also in a sample of healthy controls, with only anecdotic exceptions. As inferred from complete genome sequences available in GenBank and the literature, several of these mutations are likely to be rare variants that normally characterize mtDNA genomes in populations or even some minor haplogroups (e.g. 10394C>T is the unique diagnostic site for H16 (Álvarez-Iglesias et al. 2009; Roostalu et al. 2007)). The fact that these variants are non-synonymous or highly conserved in comparisons between species do not guaranty their disease status (Salas et al. 2009). Their rareness together with the fact that much of this variation is normally located at the tips of the phylogenies in the mtDNA tree (see e.g. [www.phylotree.org](http://www.phylotree.org)), is often interpreted as indicative of their pathogenic condition by many authors, but without any other solid support (Bandelt et al. 2006; Bandelt et al. 2009; Kong et al. 2006). The simple appearance



of a mutation in Mitomap as pathogenic has been over-interpreted (Bandelt et al. 2008). The results of the present article indicate that it is unlikely that these mutations play an important role in ASD.

We have also carried out a case-control association study in ASD patients and we could not find any evidence relating common polymorphisms with ASD. We propose three explanations for this result: (i) Type error II due to a limited sample size of cases (although statistical power is still high for the most frequent mtSNPs and hgs); (ii) mtDNA haplogroup background as analyzed in the present study is not a risk factor in ASD; and (iii) there could be a complex interplay between mtDNA variation and a nuclear factor. Although previous findings are in agreement with our results (Kent et al. 2008), we consider that further studies would be needed in order to rule out definitely any role of mtDNA haplogroups in ASD.

Summarizing, although it is widely accepted that some forms of ASD appear concomitantly with impairment of mitochondrial energy metabolism, there are reasons to believe that the cause of these mitochondrial disorders do not systematically rest on mutations or variants in the mtDNA molecule. Pathogenic mtDNA mutations have been reported in ASD patients, but this seems to be an exception rather than the rule in ASD patients. It is most likely that the real cause of mitochondrial deficiencies in ASD are due to nuclear factors acting alone or through a complex interplay with mtDNA variants.

### **Acknowledgements**

We thank the patients for their participation in the present project. This project was supported by grants from Xunta de Galicia (PGIDIT06PXIB208079PR and Grupos Emerxentes: 2008/037), Fundación de Investigación Médica Mutua Madrileña (2008/CL444) and Ministerio de Ciencia e Innovación (SAF2008-02971) given to AS.

### **References**

Álvarez-Iglesias V, Barros F, Carracedo Á, Salas A (2008) Minisequencing mitochondrial DNA pathogenic mutations. *BMC Med Genet* 9: 26.

- Álvarez-Iglesias V, Mosquera-Miguel A, Cerezo M, Quintáns B, Zarrabeitia MT, Cuscó I, Lareu MV, García O, Pérez-Jurado L, Carracedo Á, Salas A (2009) New population and phylogenetic features of the internal variation within mitochondrial DNA macro-haplogroup R0. *PLoS ONE* 4: e5112.
- Andrews RM, Kubacka I, Chinnery PF, Lightowlers RN, Turnbull DM, Howell N (1999) Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. *Nat Genet* 23: 147.
- Bandelt HJ, Salas A, Bravi CM (2006) What is a 'novel' mtDNA mutation - and does 'novelty' really matter? *J Hum Genet* 51: 1073-82.
- Bandelt HJ, Salas A, Taylor RW, Yao YG (2009) The exaggerated status of "novel" and "pathogenic" mtDNA sequence variants due to inadequate database searches. *Hum Mutat* 30: 191-6.
- Bandelt HJ, Yao YG, Salas A (2008) The search of 'novel' mtDNA mutations in hypertrophic cardiomyopathy: MITOMAPping as a risk factor. *Int J Cardiol* 126: 439-42.
- Correia C, Coutinho AM, Diogo L, Grazina M, Marques C, Miguel T, Ataíde A, Almeida J, Borges L, Oliveira C, Oliveira G, Vicente AM (2006) Brief report: High frequency of biochemical markers for mitochondrial dysfunction in autism: no association with mitochondrial aspartate/glutamate carrier *SLC25A12* gene. *J Autism dev Disord* 36: 1137-40.
- Cuscó I, Medrano A, Gener B, Vilardell M, Gallastegui F, Villa O, González E, Rodríguez-Santiago B, Vilella E, Del Campo M, Pérez-Jurado LA (2009) Autism-specific copy number variants further implicate the phosphatidylinositol signaling pathway and the glutamatergic synapse in the etiology of the disorder. *Hum Mol Genet* 18: 1795-804.
- Chugani DC, Sundram BS, Behen M, Lee ML, Moore GJ (1999) Evidence of altered energy metabolism in autistic children. *Prog Neuropsychopharmacol Biol Psychiatry* 23: 635-41.
- Filipek PA, Juranek J, Smith M, Mays LZ, Ramos ER, Bocian M, Masser-Frye D, Laulhere TM, Modahl C, Spence MA, Gargus JJ (2003) Mitochondrial dysfunction in autistic patients with 15q inverted duplication. *Ann Neurol* 53: 801-4.

- Folstein SE, Rosen-Sheidley B (2001) Genetics of autism: complex aetiology for a heterogeneous disorder. *Nat Rev Genet* 2: 943-55.
- Gauderman, W. J. 2002. Sample size requirements for matched case-control studies of gene-environment interaction. *Stat Med* 21 (1):35-50.
- Graf WD, Marin-Garcia J, Gao HG, Pizzo S, Naviaux RK, Markusic D, Barshop BA, Courchesne E, Haas RH (2000) Autism associated with the mitochondrial DNA G8363A transfer RNA(Lys) mutation. *J Child Neurol* 15: 357-61.
- Kent L, Gallagher L, Elliott HR, Mowbray C, Chinnery PF (2008) An investigation of mitochondrial haplogroups in autism. *Am J Med Genet B Neuropsychiatr Genet* 147B: 987-9.
- Kong QP, Bandelt HJ, Sun C, Yao YG, Salas A, Achilli A, Wang CY, Zhong L, Zhu CL, Wu SF, Torroni A, Zhang YP (2006) Updating the East Asian mtDNA phylogeny: a prerequisite for the identification of pathogenic mutations. *Hum Mol Genet* 15: 2076-86.
- Laszlo A, Horvath E, Eck E, Fekete M (1994) Serum serotonin, lactate and pyruvate levels in infantile autistic children. *Clin Chim Acta* 229: 205-7.
- Lombard J (1998) Autism: a mitochondrial disorder? *Med Hypotheses* 50: 497-500.
- Minshew NJ, Goldstein G, Dombrowski SM, Panchalingam K, Pettegrew JW (1993) A preliminary <sup>31</sup>P MRS study of autism: evidence for undersynthesis and increased degradation of brain membranes. *Biol Psychiatry* 33: 762-73.
- Mitchell AL, Elson JL, Howell N, Taylor RW, Turnbull DM (2006) Sequence variation in mitochondrial complex I genes: mutation or polymorphism? *J Med Genet* 43: 175-9.
- Mosquera-Miguel A, Álvarez-Iglesias V, Lareu MV, Carracedo Á, Salas A (2009) Testing the performance of mtSNP minisequencing in forensic samples. *Forensic Sci Int Genet* 3: 261-4.
- Mosquera-Miguel A, Álvarez-Iglesias V, Vega A, Milne R, Cabrera de León A, Benitez J, Carracedo Á, Salas A (2008) Is mitochondrial DNA variation associated with sporadic breast cancer risk? *Cancer Res* 68: 623-625.
- Oliveira G, Diogo L, Grazina M, Garcia P, Ataíde A, Marques C, Miguel T, Borges L, Vicente AM, Oliveira CR (2005) Mitochondrial dysfunction in autism

- spectrum disorders: a population-based study. *Dev Med Child Neurol* 47: 185-9.
- Pons R, Andreu AL, Checcarelli N, Vila MR, Engelstad K, Sue CM, Shungu D, Haggerty R, de Vivo DC, DiMauro S (2004) Mitochondrial DNA abnormalities and autistic spectrum disorders. *J Pediatr* 144: 81-5.
- Quintáns B, Álvarez-Iglesias V, Salas A, Phillips C, Lareu MV, Carracedo Á (2004) Typing of mitochondrial DNA coding region SNPs of forensic and anthropological interest using SNaPshot minisequencing. *Forensic Sci Int* 140: 251-7.
- Ramoz N, Reichert JG, Smith CJ, Silverman JM, Beshpalova IN, Davis KL, Buxbaum JD (2004) Linkage and association of the mitochondrial aspartate/glutamate carrier SLC25A12 gene with autism. *Am J Psychiatry* 161: 662-9.
- Rapin I (1999) Autism in search of a home in the brain. *Neurology* 52: 902-4.
- Roostalu U, Kutuev I, Loogväli E-L, Metspalu E, Tambets K, Reidla M, Khusnutdinova EK, Usanga E, Kivisild T, Villems R (2007) Origin and expansion of haplogroup H, the dominant human mitochondrial DNA lineage in West Eurasia: the Near Eastern and Caucasian perspective. *Mol Biol Evol* 24: 436-48.
- Salas A, Comas D, Lareu MV, Bertranpetit J, Carracedo Á (1998) mtDNA analysis of the Galician population: a genetic edge of European variation. *Eur J Hum Genet* 6: 365-75.
- Salas A, Fachal L, Marcos-Alonso S, Vega A, Martínón-Torres F, ESIGEM G (2009) Investigating the role of mitochondrial haplogroups in genetic predisposition to meningococcal disease. *PLoS One* 4: e8347.
- Veenstra-VanderWeele J, Cook EH, Jr. (2004) Molecular genetics of autism spectrum disorder. *Mol Psychiatry* 9: 819-32.
- Weissman JR, Kelley RI, Bauman ML, Cohen BH, Murray KF, Mitchell RL, Kern RL, Natowicz MR (2008) Mitochondrial disease in autism spectrum disorder patients: a cohort analysis. *PLoS One* 3: e3815.

**Table I.** Medical history and demographic data of ASD patients.

	ADULT COHORT	CHILD COHORT
<b>Sex (M:F ratio)</b>	2.1	4.75
<b>Actual Age (Mean years)</b>	42.5	10
<b>Ethnicity (%)</b>		
Western Europeans <sup>1</sup>	66 (100%)	85 (92.4%)
Other ethnicities		7 (7.6%)
<b>Epilepsy</b>	28 (42.4%)	11 (12%)
<b>Dimorphism</b>	26 (39%)	8 (8.7%)
<b>Molecular findings</b> <sup>2</sup>	5 (7.5%)	6 (6.5%)
<b>Comorbid Conditions</b> <sup>3</sup>	1 (1.5%)	5 (5.4%)
<b>Cromosomopathies</b> <sup>4</sup>	4 (6%)	-

<sup>1</sup>Western Europeans refers to people born mainly in Catalonia (northeast Spain), but some are from other spanish locations.

<sup>2</sup>Molecular finding includes 9 alterations detected by aCGH technology (Cusco et al. 2009) and 2 additional findings (SHANK3 Deletion and EHMT1 Deletion).

<sup>3</sup>Comorbid Conditions includes Fragile-X mutation (1) and pre-mutations (2); Meningitis (1) and Creatine deficiency by SLC6A8 mutation (2) (Gener et al. submitted)

<sup>4</sup>Cromosomopathies includes: 1p-ter deletions (2); Indupdel 8p (1) and Dup3q+Del11q (1) (Gener et al. submitted)

**Table II.** Pearson's chi-square test for cases *versus* controls of SNPs and most frequent haplogroups.

	rCRS ref. <sup>a</sup>	MA <sup>b</sup>	MAF <sup>c</sup>	CHI2 EXACT	P-VALUE	ADJUSTED P-VALUE <sup>d</sup>	OR (95% CI) <sup>e</sup>
<b>mtSNPs</b>							
<b>3010G&gt;A</b>	G	A	0.27	1.3239	0.250	0.9776	0.79 (0.52-1.19)
<b>3915G&gt;A</b>	G	A	0.04	1.1782	0.278	0.9872	0.52 (0.16-1.73)
<b>3992C&gt;T</b>	C	T	0.01	1.4306	0.232	0.9673	-
<b>4216T&gt;C</b>	T	C	0.16	0.7448	0.388	0.9988	0.80 (0.49-1.32)
<b>4336T&gt;C</b>	T	C	0.04	0.0006	0.980	1	1.01 (0.41-2.49)
<b>4529A&gt;T</b>	A	T	0.01	0.0420	0.838	1	1.18 (0.25-5.60)
<b>4580G&gt;A</b>	G	A	0.03	2.0152	0.156	0.9084	1.75 (0.80-3.83)
<b>4769A&gt;G</b>	A	A	0.02	3.8245	0.051	0.5133	-
<b>4793A&gt;C</b>	A	G	0.01	1.2709	0.260	0.9817	-
<b>6776T&gt;C</b>	T	C	0.05	0.8820	0.348	0.9963	0.64 (0.25-1.65)
<b>7028C&gt;T</b>	C	T	0.44	1.8561	0.173	0.9260	1.28 (0.90-1.82)
<b>10398A&gt;G</b>	A	G	0.20	0.8362	0.360	0.9978	0.81 (0.51-1.28)
<b>10400C&gt;T</b>	C	T	0.01	0.7699	0.380	0.9987	1.81 (0.47-6.88)
<b>10463T&gt;C</b>	T	C	0.09	0.0002	0.987	1	1.01 (0.54-1.88)
<b>10873T&gt;C</b>	C	C	0.04	2.1633	0.141	0.8728	0.58 (0.27-1.21)
<b>12308A&gt;G</b>	A	G	0.23	1.2282	0.268	0.9843	0.78 (0.51-1.21)
<b>12705C&gt;T</b>	C	T	0.08	5.7751	0.016	0.1933	1.88 (1.12-3.18)
<b>13966A&gt;G</b>	A	G	0.02	0.0103	0.919	1	0.94 (0.27-3.28)
<b>14766C&gt;T</b>	C	C	0.49	0.0618	0.804	1	0.96 (0.68-1.35)
<b>Haplogroups</b>							
<b>H</b>	-	-	0.44	1.5221	0.217	0.9572	1.25 (0.88-1.77)
<b>H1</b>	-	-	0.20	1.6893	0.194	0.9442	1.36 (0.85-2.18)
<b>HV</b>	-	-	0.47	0.0801	0.777	1	1.05 (0.75-1.48)
<b>UK</b>	-	-	0.23	1.2796	0.258	0.9806	1.28 (0.83-1.98)
<b>J</b>	-	-	0.08	2.9110	0.088	0.7090	1.98 (0.89-4.41)
<b>T</b>	-	-	0.08	0.0010	0.975	1	0.99 (0.53-1.85)
<b>JT</b>	-	-	0.16	1.4932	0.222	0.9590	1.37 (0.83-2.27)

<sup>a</sup>rCRS: allele in the revised Cambridge Reference Sequence (rCRS) (Andrews et al. 1999);

<sup>b</sup>MA: allele with the lowest frequency;

<sup>c</sup>MAF: minimum allele frequency computed on control individuals;

<sup>d</sup>Adjusted *P*-value: adjustment of chi-square *P*-values was carried out with a permutation-based approach; number of permutations = 20,000;

<sup>e</sup>OR (95%CI): ORs were computed with the rCRS allele as a reference.

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	AA	AB	AC	AD	AE	AF	AG	AH	AI	AJ	AK	AL	AM	AN	AO	AP	AQ	AR	AS	AT	AU	AV	AW	AX	AY	AZ	BA	BB	BC	BD	BE	BF	BG	BH	BI	BJ	BK	BL	BM	BN	BO	BP	BQ	BR	BS	BT	BU	BV	BW	BX	BY	BZ	CA	CB	CC	CD	CE	CF	CG	CH	CI	CJ	CK	CL	CM	CN	CO	CP	CQ	CR	CS	CT	CU	CV	CW	CX	CY	CZ	DA	DB	DC	DD	DE	DF	DG	DH	DI	DJ	DK	DL	DM	DN	DO	DP	DQ	DR	DS	DT	DU	DV	DW	DX	DY	DZ	EA	EB	EC	ED	EE	EF	EG	EH	EI	EJ	EK	EL	EM	EN	EO	EP	EQ	ER	ES	ET	EU	EV	EW	EX	EY	EZ	FA	FB	FC	FD	FE	FF	FG	FH	FI	FJ	FK	FL	FM	FN	FO	FP	FQ	FR	FS	FT	FU	FV	FW	FX	FY	FZ	GA	GB	GC	GD	GE	GF	GG	GH	GI	GJ	GK	GL	GM	GN	GO	GP	GQ	GR	GS	GT	GU	GV	GW	GX	GY	GZ	HA	HB	HC	HD	HE	HF	HG	HH	HI	HJ	HK	HL	HM	HN	HO	HP	HQ	HR	HS	HT	HU	HV	HW	HX	HY	HZ	IA	IB	IC	ID	IE	IF	IG	IH	II	IJ	IK	IL	IM	IN	IO	IP	IQ	IR	IS	IT	IU	IV	IW	IX	IY	IZ	JA	JB	JC	JD	JE	JF	JG	JH	JI	JJ	JK	JL	JM	JN	JO	JP	JQ	JR	JS	JT	JU	JV	JW	JX	JY	JZ	KA	KB	KC	KD	KE	KF	KG	KH	KI	KJ	KK	KL	KM	KN	KO	KP	KQ	KR	KS	KT	KU	KV	KW	KX	KY	KZ	LA	LB	LC	LD	LE	LF	LG	LH	LI	LJ	LK	LL	LM	LN	LO	LP	LQ	LR	LS	LT	LU	LV	LW	LX	LY	LZ	MA	MB	MC	MD	ME	MF	MG	MH	MI	MJ	MK	ML	MM	MN	MO	MP	MQ	MR	MS	MT	MU	MV	MW	MX	MY	MZ	NA	NB	NC	ND	NE	NF	NG	NH	NI	NJ	NK	NL	NM	NN	NO	NP	NQ	NR	NS	NT	NU	NV	NW	NX	NY	NZ	OA	OB	OC	OD	OE	OF	OG	OH	OI	OJ	OK	OL	OM	ON	OO	OP	OQ	OR	OS	OT	OU	OV	OW	OX	OY	OZ	PA	PB	PC	PD	PE	PF	PG	PH	PI	PJ	PK	PL	PM	PN	PO	PP	PQ	PR	PS	PT	PU	PV	PW	PX	PY	PZ	QA	QB	QC	QD	QE	QF	QG	QH	QI	QJ	QK	QL	QM	QN	QO	QP	QQ	QR	QS	QT	QU	QV	QW	QX	QY	QZ	RA	RB	RC	RD	RE	RF	RG	RH	RI	RJ	RK	RL	RM	RN	RO	RP	RQ	RR	RS	RT	RU	RV	RW	RX	RY	RZ	SA	SB	SC	SD	SE	SF	SG	SH	SI	SJ	SK	SL	SM	SN	SO	SP	SQ	SR	SS	ST	SU	SV	SW	SX	SY	SZ	TA	TB	TC	TD	TE	TF	TG	TH	TI	TJ	TK	TL	TM	TN	TO	TP	TQ	TR	TS	TU	TV	TW	TX	TY	TZ	UA	UB	UC	UD	UE	UF	UG	UH	UI	UJ	UK	UL	UM	UN	UO	UP	UQ	UR	US	UT	UU	UV	UW	UX	UY	UZ	VA	VB	VC	VD	VE	VF	VG	VH	VI	VJ	VK	VL	VM	VN	VO	VP	VQ	VR	VS	VT	VU	VV	VW	VX	VY	VZ	WA	WB	WC	WD	WE	WF	WG	WH	WI	WJ	WK	WL	WM	WN	WO	WP	WQ	WR	WS	WT	WU	WV	WW	WX	WY	WZ	XA	XB	XC	XD	XE	XF	XG	XH	XI	XJ	XK	XL	XM	XN	XO	XP	XQ	XR	XS	XT	XU	XV	XW	XX	XY	XZ	YA	YB	YC	YD	YE	YF	YG	YH	YI	YJ	YK	YL	YM	YN	YO	YP	YQ	YR	YS	YT	YU	YV	YW	YX	YY	YZ	ZA	ZB	ZC	ZD
---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----

---

*DISCUSIÓN*





## El ADNmt en los estudios de asociación

Los estándares de calidad de los estudios de asociación entre el ADNmt y enfermedad compleja son habitualmente deficientes, ya que suelen presentar carencias en algunos de los siguientes aspectos:

- ❖ Un **muestreo deficiente** puede provocar la aparición de falsas asociaciones positivas. Los problemas más habituales relacionados con el muestreo son:
  - *Tamaño muestral.* Se requieren grandes cohortes para aumentar el poder de detección de una posible asociación entre el ADNmt y una enfermedad compleja (Samuels et al. 2006), sin embargo esto podría tener su contrapartida ya que también infla el efecto de ruido o confusión (“*confounding*”) si existe estratificación poblacional.
  - *Clasificación étnica basada en el fenotipo.* El fenotipo no tiene porque ser *a priori* un buen predictor del linaje mitocondrial de las muestras (Salas et al. 2007), y la inclusión de un pequeño número de individuos con un fondo genético diferente puede ser un factor de riesgo frecuente para errores tipo I.
  - *Estratificación poblacional.* Este fenómeno provoca la ausencia de identidad entre casos y controles, y la consiguiente aparición de asociaciones espurias. El riesgo de falsos positivos es particularmente alto en los estudios de ADNmt, ya que su variabilidad está más profundamente estructurada en poblaciones humanas debido al menor tamaño poblacional efectivo en comparación con marcadores autosómicos (Kong et al. 2006). El control de este fenómeno es delicado ya que los métodos empleados para controlar y monitorizar la estratificación poblacional en cromosomas autosómicos (ej. control genómico; (Cardon and Palmer 2003)) no son probablemente adecuados en un contexto mitocondrial.
- ❖ Son escasos los estudios en los que se lleva a cabo una revisión adecuada desde una **perspectiva filogenética** de los datos obtenidos. En la actualidad ya existe una cantidad considerable de información en relación a la filogenia mitocondrial basada en la secuenciación de genomas completos. Se conocen con cierto detalle tanto el esqueleto principal como las ramas más jóvenes de

la filogenia mitocondrial, por lo que cada nuevo perfil de ADNmt debe situarse en alguna rama del árbol filogenético (Bandelt et al. 2005a; Kong et al. 2006).

- Es siempre recomendable comprobar que la supuesta mutación “novel” no forma parte de la colección de polimorfismos que caracterizan un haplogrupo mitocondrial antes de formular conclusiones precipitadas. Cuando un perfil no se ajusta completamente a la filogenia, puede representar una rama basal todavía no observada, o puede tratarse de una secuencia con artefactos (contaminación cruzada de muestras).
  - El análisis de los datos desde un punto de vista filogenético es un control interno que permite detectar artefactos de genotipado o mezcla de muestras.
  - Es fundamental referir las variantes con respecto a la rCRS; una lista incompleta es de poca ayuda a la hora de entender las bases genéticas de una enfermedad.
  - Aquellas posiciones polimórficas situadas en ramas antiguas del árbol filogenético han estado bajo presión selectiva miles de años, por lo que su amplia distribución no sugiere directamente la posibilidad de un efecto patogénico.
  - *Mal planteamiento de hipótesis.* El ADNmt no experimenta recombinación, de forma que los polimorfismos se heredan como un haplotipo, y por tanto no son alelos independientes. Por ello, no se pueden tratar los SNPs del ADNmt como alelos y cualquier análisis de interacción (*sensu* epistasia) no sería más que un análisis de haplogrupos/haplotipos.
- ❖ **Réplica de los resultados.** La asociación de un mtSNP con una enfermedad, debe ser replicada empleando otra muestra recogida de manera independiente (que represente al mismo grupo poblacional), ya que si realmente posee algún efecto en el riesgo a padecer una enfermedad debería detectarse nuevamente. La ausencia de réplicas en los estudios de asociación y la disparidad en los resultados hacen que las conclusiones sean inverosímiles.
- ❖ **Análisis estadísticos deficientes.** En este tipo de estudios es relativamente frecuente detectar ausencia de correcciones por test múltiple, y no existe una preocupación excesiva por el control de la estratificación poblacional.

- ❖ **Artefactos en el genotipado.** Cuando el genotipado se basa en la secuenciación, es imprescindible asegurarse de que no se trata de una mutación fantasma (Brandstätter et al. 2005), por lo que debe ser detectada en las dos direcciones de lectura, *forward* y *reverse*. Si los datos se obtienen mediante el empleo de enzimas de restricción (RFLPs) hay que tener presente que, aún siendo la muestra homoplásmica, existe un riesgo alto de digestión parcial del ADN, obteniendo así perfiles incompletos que pueden ser interpretados como heteroplasmías (Bandelt et al. 2005b). Sesgos o problemas en el genotipado de SNPs pueden también contribuir a falsos positivos en estudios de asociación del ADNmt (Salas et al. 2009).
- ❖ **Ausencia de estudios funcionales.** La formación de cíbridos, por ejemplo, permitiría confirmar que la mutación asociada con la enfermedad tiene un efecto en la disfunción mitocondrial. Se debe advertir sin embargo que los problemas que pueden generarse en el estudio de cíbridos aun no han sido estudiados con detalle.
- ❖ **Criterios de patogenicidad.** Es necesario tener en cuenta criterios que faciliten la identificación de nuevas mutaciones patogénicas (Dimauro and Davidzon 2005; McFarland et al. 2004; Mitchell et al. 2006), ya que por sí solo el estado “no sinónimo” de una variante no es suficiente para determinar su patogenicidad porque muchos SNPs de la filogenia mitocondrial suponen cambios no sinónimos. Por otra parte, también es imprescindible realizar una revisión intensiva de la literatura (varias bases de datos y buscadores menos específicos) antes de calificar una mutación como novel (Bandelt et al. 2009).
- ❖ **NUMTs.** Se ha sugerido que la detección de NUMTs podría estar incorporando ruido a los resultados (heteroplasmías) (Yao et al. 2008), por lo que es aconsejable emplear protocolos que eviten en la medida de lo posible el genotipado de estas regiones nucleares (Ramos et al. 2009).

## El ADNmt en genética de poblaciones y forense

Tanto desde el punto de vista de la genética clínica como en el campo forense, es importante conocer adecuadamente la composición genética de las poblaciones de estudio, ya que ello permitirá interpretar correctamente los resultados que se

obtingan. La genética de poblaciones se encarga del estudio de la variabilidad genética inter e intrapoblacional, de los movimientos migratorios que han dado lugar a las poblaciones actuales y de datar los diferentes linajes que se observan hoy en día. Es importante caracterizar bien las poblaciones y tener en cuenta que a pesar de su aparente homogeneidad en Europa, existen diferencias sustanciales entre regiones geográficas relativamente pequeñas. Otros aspectos de interés en el estudio del genoma mitocondrial están relacionados con el conocimiento de la filogenia en linajes concretos, lo cual es de vital importancia para el diseño de paneles de SNPs de discriminación poblacional, o la detección de errores generados *in vitro* (mutaciones fantasma, recombinación artificial, etc), conocimiento de las tasas de mutación, etc.

No cabe duda, de que la pericia forense es una de las más beneficiadas del conocimiento adquirido en genética de poblaciones. A pesar de que el genoma mitocondrial no sea uno de los *tests* de rutina, sí es importante en determinados contextos como es el caso de pelos sin bulbo o muestras muy degradadas. En este sentido, el conocimiento de la composición genética de la población a la que pertenece una muestra forense o las tasas sitio-específicas de las posiciones *hotspots* del ADNmt son esenciales para poder interpretar adecuadamente el peso de la prueba. El desarrollo de estrategias que mejoren el éxito del análisis de muestras degradadas o de baja concentración de ADN, como el genotipado de mtSNPs, también es necesario para mejorar la discriminación de los perfiles de ADNmt.

*CONCLUSIONES*



## Conclusiones de los estudios de asociación caso-control

1. No se encontraron evidencias de asociación entre ninguna de las 25 variantes comunes del ADNmt genotipadas y el riesgo de padecer cáncer de mama esporádico, tras corregir mediante test múltiple, para ninguna de las dos muestras españolas estudiadas (península Ibérica e islas Canarias).
2. En el estudio de asociación de enfermedad meningocócica, el empleo de dos grupos de controles diferentes (CG1 y CG2) permitió eliminar la presunta asociación entre el polimorfismo G11719A y la EM. La causa más probable para explicar la ambigüedad de estos resultados es el efecto de estratificación poblacional residual que no pudo ser eliminada mediante las herramientas empleadas. Problemas en el genotipado también pueden provocar la aparición de falsos positivos.
3. En contra de los estudios publicados hasta la fecha, nuestros resultados derivados del análisis de la mayor cohorte de pacientes esquizofrénicos analizada hasta la fecha, sugieren que los principales linajes mitocondriales europeos no son un factor de riesgo a padecer esquizofrenia. Tampoco se ha detectado asociación entre la aparición de los síntomas y linajes mitocondriales.
4. En relación al estudio de asociación caso-control entre la variabilidad del ADNmt y los trastornos del espectro autista (ASD), no se han detectado diferencias significativas en la distribución de haplogrupos de ambas cohortes. Tampoco se ha detectado ninguna de las mutaciones patogénicas confirmadas como causantes de enfermedades mitocondriales (Álvarez-Iglesias et al. 2008). Se encontraron tres variantes “potencialmente” patogénicas, sin embargo, dos de ellas fueron detectadas en el grupo control y una en el grupo de casos. En conclusión, no se ha detectado asociación significativa entre la variabilidad del ADNmt como factor de riesgo a los ASD.
5. La demostración de una presunta asociación entre la variabilidad del ADNmt y la susceptibilidad, curso o pronóstico de determinadas enfermedades complejas como el cáncer de mama, esquizofrenia, enfermedad meningocócica o autismo, debe hacerse sobre la base de un buen diseño

experimental, que implique por lo menos un tamaño muestral adecuado, una correcta selección de casos y controles, y un estudio de réplica que garantice en cierta medida los resultados obtenidos con una primera cohorte.

6. Existen grandes deficiencias en el tratamiento estadístico en los estudios de asociación, que junto con otros problemas metodológicos están creando serias dudas sobre la presunta asociación de variantes mitocondriales con la enfermedad compleja.

## Conclusiones de los estudios genético poblacionales y forenses

1. El estudio poblacional llevado a cabo revela niveles moderados de estratificación en el norte de la península Ibérica. Galicia y Cantabria presentan valores de diversidad bajos, especialmente para el haplogrupo H.
2. Se han observado patrones geográficos, estadísticamente clinales, para algunos linajes mitocondriales (H1, H3 y H5a) con picos de frecuencia en la región Franco-cantábrica disminuyendo hacia Europa del este, compatibles con la repoblación de Europa después del Último Máximo Glacial (LGM, *Late Glacial Maximum*).
3. El conjunto de 71 SNPs seleccionados para cubrir la filogenia de R0, y que aumentan considerablemente el poder de discriminación del *test* mitocondrial en muestras europeas, funciona de forma adecuada para el genotipado de muestras forenses degradadas.
4. Las técnicas de minisecuenciación son altamente reproducibles y robustas ya que hasta el momento, no hemos podido detectar falsos positivos o artefactos de genotipado que interfirieran con la lectura apropiada del perfil mitocondrial, así como ningún patrón indicativo de contaminación de muestras ni de amplificación involuntaria de NUMTs.



---

## *BIBLIOGRAFÍA*



- Achilli A, Perego UA, Bravi CM, Coble MD, Kong Q-P, Woodward SR, Salas A, Torroni A, Bandelt H-J (2008) The phylogeny of the four pan-American MtDNA haplogroups: implications for evolutionary and disease studies. *PLoS ONE* 3: e1764
- Álvarez-Iglesias V, Barros F, Carracedo A, Salas A (2008) Minisequencing mitochondrial DNA pathogenic mutations. *BMC Med Genet* 9: 26
- Álvarez-Iglesias V, Jaime JC, Carracedo A, Salas A (2007) Coding region mitochondrial DNA SNPs: targeting East Asian and Native American haplogroups. *Forensic Sci Int Genet* 1: 44-55
- Álvarez-Iglesias V, Mosquera-Miguel A, Cerezo M, Quintáns B, Zarrabeitia MT, Cusco I, Lareu MV, García O, Pérez-Jurado L, Carracedo A, Salas A (2009) New population and phylogenetic features of the internal variation within mitochondrial DNA macro-haplogroup R0. *PLoS ONE* 4: e5112
- Amar S, Shamir A, Ovadia O, Blanaru M, Reshef A, Kremer I, Rietschel M, Schulze TG, Maier W, Belmaker RH, Ebstein RP, Agam G, Mishmar D (2007) Mitochondrial DNA HV lineage increases the susceptibility to schizophrenia among Israeli Arabs. *Schizophr Res* 94: 354-8
- Amo T, Yadava N, Oh R, Nicholls DG, Brand MD (2008) Experimental assessment of bioenergetic differences caused by the common European mitochondrial DNA haplogroups H and T. *Gene* 411: 69-76
- Anderson S, Bankier AT, Barrell BG, de Bruijn MH, Coulson AR, Drouin J, Eperon IC, Nierlich DP, Roe BA, Sanger F, Schreier PH, Smith AJ, Staden R, Young IG (1981) Sequence and organization of the human mitochondrial genome. *Nature* 290: 457-65
- Andrews RM, Kubacka I, Chinnery PF, Lightowlers RN, Turnbull DM, Howell N (1999) Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. *Nat Genet* 23: 147
- Awadalla P, Eyre-Walker A, Smith JM (1999) Linkage disequilibrium and recombination in hominid mitochondrial DNA. *Science* 286: 2524-5
- Bai RK, Leal SM, Covarrubias D, Liu A, Wong LJ (2007) Mitochondrial genetic background modifies breast cancer risk. *Cancer Res* 67: 4687-94
- Bamne MN, Talkowski ME, Moraes CT, Manuck SB, Ferrell RE, Chowdari KV, Nimgaonkar VL (2008) Systematic association studies of mitochondrial

- DNA variations in schizophrenia: focus on the ND5 gene. *Schizophr Bull* 34: 458-65
- Bandelt H-J, Achilli A, Kong Q-P, Salas A, Lutz-Bonengel S, Sun C, Zhang Y-P, Torroni A, Yao Y-G (2005a) Low "penetrance" of phylogenetic knowledge in mitochondrial disease studies. *Biochem Biophys Res Commun* 333: 122-30
- Bandelt H-J, Kong Q-P, Parson W, Salas A (2005b) More evidence for non-maternal inheritance of mitochondrial DNA? *J Med Genet* 42: 957-60
- Bandelt H-J, Olivieri A, Bravi C, Yao Y-G, Torroni A, Salas A (2007) 'Distorted' mitochondrial DNA sequences in schizophrenic patients. *Eur J Hum Genet* 15: 400-2; author reply 402-4
- Bandelt H-J, Parson W (2008) Consistent treatment of length variants in the human mtDNA control region: a reappraisal. *Int J Legal Med* 122: 11-21
- Bandelt H-J, Salas A, Taylor RW, Yao Y-G (2009) Exaggerated status of "novel" and "pathogenic" mtDNA sequence variants due to inadequate database searches. *Hum Mutat* 30: 191-6
- Bandelt H-J, Yao Y-G, Kivisild T (2005c) Mitochondrial genes and schizophrenia. *Schizophr Res* 72: 267-9
- Barros F, Lareu MV, Salas A, Carracedo A (1997) Rapid and enhanced detection of mitochondrial DNA variation using single-strand conformation analysis of superposed restriction enzyme fragments from polymerase chain reaction-amplified products. *Electrophoresis* 18: 52-4
- Baudouin SV, Saunders D, Tiangyou W, Elson JL, Poynter J, Pyle A, Keers S, Turnbull DM, Howell N, Chinnery PF (2005) Mitochondrial DNA and survival after sepsis: a prospective study. *Lancet* 366: 2118-21
- Ben-Shachar D (2002) Mitochondrial dysfunction in schizophrenia: a possible linkage to dopamine. *J Neurochem* 83: 1241-51
- Bendall KE, Macaulay VA, Baker JR, Sykes BC (1996) Heteroplasmic point mutations in the human mtDNA control region. *Am J Hum Genet* 59: 1276-87
- Bodenteich A, Mitchell LG, Polymeropoulos MH, Merril CR (1992) Dinucleotide repeat in the human mitochondrial D-loop. *Hum Mol Genet* 1: 140

- Brandstätter A, Parsons TJ, Parson W (2003) Rapid screening of mtDNA coding region SNPs for the identification of west European Caucasian haplogroups. *Int J Legal Med* 117: 291-8
- Brandstätter A, Peterson CT, Irwin JA, Mpoke S, Koech DK, Parson W, Parsons TJ (2004) Mitochondrial DNA control region sequences from Nairobi (Kenya): inferring phylogenetic parameters for the establishment of a forensic database. *Int J Legal Med* 118: 294-306
- Brandstätter A, Sanger T, Lutz-Bonengel S, Parson W, Beraud-Colomb E, Wen B, Kong QP, Bravi CM, Bandelt H-J (2005) Phantom mutation hotspots in human mitochondrial DNA. *Electrophoresis* 26: 3414-29
- Brown WM, George M, Jr., Wilson AC (1979) Rapid evolution of animal mitochondrial DNA. *Proc Natl Acad Sci U S A* 76: 1967-71
- Calloway CD, Reynolds RL, Herrin GL, Jr., Anderson WW (2000) The frequency of heteroplasmy in the HVII region of mtDNA differs across tissue types and increases with age. *Am J Hum Genet* 66: 1384-97
- Cann RL, Stoneking M, Wilson AC (1987) Mitochondrial DNA and human evolution. *Nature* 325: 31-6
- Canter JA, Kallianpur AR, Parl FF, Millikan RC (2005) Mitochondrial DNA G10398A polymorphism and invasive breast cancer in African-American women. *Cancer Res* 65: 8028-33
- Cardon LR, Palmer LJ (2003) Population stratification and spurious allelic association. *Lancet* 361: 598-604
- Carew JS, Huang P (2002) Mitochondrial defects in cancer. *Mol Cancer* 1: 9
- Carracedo A, Bär W, Lincoln P, Mayr W, Morling N, Olaisen B, Schneider P, Budowle B, Brinkmann B, Gill P, Holland M, Tully G, Wilson M (2000) DNA commission of the international society for forensic genetics: guidelines for mitochondrial DNA typing. *Forensic Sci Int* 110: 79-85
- Coble MD, Loreille OM, Wadhams MJ, Edson SM, Maynard K, Meyer CE, Niederstatter H, Berger C, Berger B, Falsetti AB, Gill P, Parson W, Finelli LN (2009) Mystery solved: the identification of the two missing Romanov children using DNA analysis. *PLoS One* 4: e4838
- Costas J, Carrera N, Dominguez E, Vilella E, Martorell L, Valero J, Gutierrez-Zotes A, Labad A, Carracedo A (2009) A common haplotype of DRD3 affected by

- recent positive selection is associated with protection from schizophrenia. *Hum Genet* 124: 607-13
- Covarrubias D, Bai RK, Wong LJ, Leal SM (2008) Mitochondrial DNA variant interactions modify breast cancer risk. *J Hum Genet* 53: 924-8
- Czarnecka AM, Krawczyk T, Zdrozny M, Lubinski J, Arnold RS, Kukwa W, Scinska A, Golik P, Bartnik E, Petros JA (2009) Mitochondrial NADH-dehydrogenase subunit 3 (ND3) polymorphism (A10398G) and sporadic breast cancer in Poland. *Breast Cancer Res Treat*
- Chen Y-S, Torroni A, Excoffier L, Santachiara-Benerecetti AS, Wallace DC (1995) Analysis of mtDNA variation in African populations reveals the most ancient of all human continent-specific haplogroups. *Am J Hum Genet* 57: 133-49
- Chinnery PF, Taylor GA, Howell N, Andrews RM, Morris CM, Taylor RW, McKeith IG, Perry RH, Edwardson JA, Turnbull DM (2000) Mitochondrial DNA haplogroups and susceptibility to AD and dementia with Lewy bodies. *Neurology* 55: 302-4
- Denaro M, Blanc H, Johnson MJ, Chen KH, Wilmsen E, Cavalli-Sforza LL, Wallace DC (1981) Ethnic variation in Hpa 1 endonuclease cleavage patterns of human mitochondrial DNA. *Proc Natl Acad Sci U S A* 78: 5768-72
- DiMauro S (2004) Mitochondrial diseases. *Biochim Biophys Acta* 1658: 80-8
- Dimauro S, Davidzon G (2005) Mitochondrial DNA and disease. *Ann Med* 37: 222-32
- Doi N, Hoshi Y (2007) Persistence problem in schizophrenia and mitochondrial DNA. *Am J Med Genet B Neuropsychiatr Genet* 144B: 1-4
- Dyall SD, Brown MT, Johnson PJ (2004) Ancient invasions: from endosymbionts to organelles. *Science* 304: 253-7
- Elson JL, Andrews RM, Chinnery PF, Lightowlers RN, Turnbull DM, Howell N (2001) Analysis of European mtDNAs for recombination. *Am J Hum Genet* 68: 145-153
- Elson JL, Turnbull DM, Taylor RW (2007) Testing the adaptive selection of human mtDNA haplogroups: an experimental bioenergetics approach. *Biochem J* 404: e3-5
- Eyre-Walker A, Smith NH, Smith JM (1999) How clonal are human mitochondria? *Proc Biol Sci* 266: 477-83

- Filosto M, Mancuso M, Vives-Bauza C, Vila MR, Shanske S, Hirano M, Andreu AL, DiMauro S (2003) Lack of paternal inheritance of muscle mitochondrial DNA in sporadic mitochondrial myopathies. *Ann Neurol* 54: 524-6
- Finnilä S, Hassinen IE, Ala-Kokko L, Majamaa K (2000) Phylogenetic network of the mtDNA haplogroup U in Northern Finland based on sequence analysis of the complete coding region by conformation-sensitive gel electrophoresis. *Am J Hum Genet* 66: 1017-26
- Finnilä S, Lehtonen MS, Majamaa K (2001) Phylogenetic network for European mtDNA. *Am J Hum Genet* 68: 1475-84
- Forster P, Harding R, Torroni A, Bandelt HJ (1996) Origin and evolution of Native American mtDNA variation: a reappraisal. *Am J Hum Genet* 59: 935-45
- Giles RE, Blanc H, Cann HM, Wallace DC (1980) Maternal inheritance of human mitochondrial DNA. *Proc Natl Acad Sci U S A* 77: 6715-9
- Gill P, Ivanov PL, Kimpton C, Piercy R, Benson N, Tully G, Evett I, Hagelberg E, Sullivan K (1994) Identification of the remains of the Romanov family by DNA analysis. *Nat Genet* 6: 130-5
- Gómez R, O'Keeffe T, Chang LY, Huebinger RM, Minei JP, Barber RC (2009) Association of mitochondrial allele 4216C with increased risk for complicated sepsis and death after traumatic injury. *J Trauma* 66: 850-7; discussion 857-8
- Graf WD, Marin-Garcia J, Gao HG, Pizzo S, Naviaux RK, Markusic D, Barshop BA, Courchesne E, Haas RH (2000) Autism associated with the mitochondrial DNA G8363A transfer RNA(Lys) mutation. *J Child Neurol* 15: 357-61
- Gray MW, Burger G, Lang BF (1999) Mitochondrial evolution. *Science* 283: 1476-81
- Greenberg BD, Newbold JE, Sugino A (1983) Intraspecific nucleotide sequence variability surrounding the origin of replication in human mitochondrial DNA. *Gene* 21: 33-49
- Hagelberg E (2003) Recombination or mutation rate heterogeneity? Implications for Mitochondrial Eve. *Trends Genet* 19: 84-90
- Hagelberg E, Goldman N, Lio P, Whelan S, Schiefenhover W, Clegg JB, Bowden DK (1999) Evidence for mitochondrial DNA recombination in a human population of island Melanesia. *Proc Biol Sci* 266: 485-92

- Hagelberg E, Goldman N, Lio P, Whelan S, Schiefenhövel W, Clegg JB, Bowden DK (2000) Evidence for mitochondrial DNA recombination in a human population of island Melanesia: correction. *Proc Biol Sci* 267: 1595-1596
- Hendrickson SL, Hutcheson HB, Ruiz-Pesini E, Poole JC, Lautenberger J, Sezgin E, Kingsley L, Goedert JJ, Vlahov D, Donfield S, Wallace DC, O'Brien SJ (2008) Mitochondrial DNA haplogroups influence AIDS progression. *AIDS* 22: 2429-39
- Herrnstadt C, Elson JL, Fahy E, Preston G, Turnbull DM, Anderson C, Ghosh SS, Olefsky JM, Beal MF, Davis RE, Howell N (2002) Reduced-median-network analysis of complete mitochondrial DNA coding-region sequences for the major African, Asian, and European haplogroups. *Am J Hum Genet* 70: 1152-71
- Herrnstadt C, Howell N (2004) An evolutionary perspective on pathogenic mtDNA mutations: haplogroup associations of clinical disorders. *Mitochondrion* 4: 791-8
- Heyer E, Zietkiewicz E, Rochowski A, Yotova V, Puymirat J, Labuda D (2001) Phylogenetic and familial estimates of mitochondrial substitution rates: study of control region mutations in deep-rooting pedigrees. *Am J Hum Genet* 69: 1113-26
- Houshmand M, Banoei MM, Tabarsi P, Panahi MS, Kashani BH, Ebrahimi G, Zargar L, Farnia P, Morris MW, Mansouri D, Velayati AA, Mirsaeidi MS (2007) Do mitochondrial DNA haplogroups play a role in susceptibility to tuberculosis? *Respirology* 12: 823-7
- Howell N, Kubacka I, Mackey DA (1996) How rapidly does the human mitochondrial genome evolve? *Am J Hum Genet* 59: 501-9
- Howell N, Smejkal CB, Mackey DA, Chinnery PF, Turnbull DM, Herrnstadt C (2003) The pedigree rate of sequence divergence in the human mitochondrial genome: there is a difference between phylogenetic and pedigree rates. *Am J Hum Genet* 72: 659-70
- Ingman M, Gyllensten U (2003) Mitochondrial genome variation and evolutionary history of Australian and New Guinean aborigines. *Genome Res* 13: 1600-6
- Ingman M, Kaessmann H, Paabo S, Gyllensten U (2000) Mitochondrial genome variation and the origin of modern humans. *Nature* 408: 708-13



- Irwin JA, Saunier JL, Niederstatter H, Strouss KM, Sturk KA, Diegoli TM, Brandstatter A, Parson W, Parsons TJ (2009) Investigation of heteroplasmy in the human mitochondrial DNA control region: a synthesis of observations from more than 5000 global population samples. *J Mol Evol* 68: 516-27
- Jeffreys AJ, Wilson V, Thein SL (1985) Individual-specific 'fingerprints' of human DNA. *Nature* 316: 76-9
- Kent L, Gallagher L, Elliott HR, Mowbray C, Chinnery PF (2008) An investigation of mitochondrial haplogroups in autism. *Am J Med Genet B Neuropsychiatr Genet* 147B: 987-9
- Kivisild T, Shen P, Wall DP, Do B, Sung R, Davis K, Passarino G, Underhill PA, Scharfe C, Torroni A, Scozzari R, Modiano D, Coppa A, de Knijff P, Feldman M, Cavalli-Sforza LL, Oefner PJ (2006) The role of selection in the evolution of human mitochondrial genomes. *Genetics* 172: 373-87
- Kivisild T, Villems R, Jorde LB, Bamshad M, Kumar S, Hedrick P, Dowling T, Stoneking M, Parsons TJ, Irwin JA, Awadalla P, Eyre-Walker A, Smith JM (2000) Questioning Evidence for Recombination in Human Mitochondrial DNA. *Science* 288: 1931a
- Kong Q-P, Bandelt H-J, Sun C, Yao Y-G, Salas A, Achilli A, Wang CY, Zhong L, Zhu CL, Wu SF, Torroni A, Zhang Y-P (2006) Updating the East Asian mtDNA phylogeny: a prerequisite for the identification of pathogenic mutations. *Hum Mol Genet* 15: 2076-86
- Kraytsberg Y, Schwartz M, Brown TA, Ebraldise K, Kunz WS, Clayton DA, Vissing J, Khrapko K (2004) Recombination of human mitochondrial DNA. *Science* 304: 981
- Luft R, Ikkos D, Palmieri G, Ernster L, Afzelius B (1962) A case of severe hypermetabolism of nonthyroid origin with a defect in the maintenance of mitochondrial respiratory control: a correlated clinical, biochemical, and morphological study. *J Clin Invest* 41: 1776-804
- Maca-Meyer N, Gonzalez AM, Larruga JM, Flores C, Cabrera VM (2001) Major genomic mitochondrial lineages delineate early human expansions. *BMC Genet* 2: 13
- Macaulay V, Hill C, Achilli A, Rengo C, Clarke D, Meehan W, Blackburn J, Semino O, Scozzari R, Cruciani F, Taha A, Shaari NK, Raja JM, Ismail P, Zainuddin Z,

- Goodwin W, Bulbeck D, Bandelt H-J, Oppenheimer S, Torroni A, Richards M (2005) Single, rapid coastal settlement of Asia revealed by analysis of complete mitochondrial genomes. *Science* 308: 1034-6
- Macaulay V, Richards M, Sykes B (1999) Mitochondrial DNA recombination-no need to panic. *Proc Biol Sci* 266: 2037-9; discussion 2041-2
- Macaulay VA, Richards MB, Forster P, Bendall KE, Watson E, Sykes B, Bandelt H-J (1997) mtDNA mutation rates--no need to panic. *Am J Hum Genet* 61: 983-90
- Magri C, Gardella R, Barlati SD, Valsecchi P, Sacchetti E, Barlati S (2007) Mitochondrial DNA haplogroups and age at onset of schizophrenia. *Am J Med Genet B Neuropsychiatr Genet* 144B: 496-501
- Malka F, Lombes A, Rojo M (2006) Organization, dynamics and transmission of mitochondrial DNA: focus on vertebrate nucleoids. *Biochim Biophys Acta* 1763: 463-72
- Marchbanks RM, Ryan M, Day IN, Owen M, McGuffin P, Whatley SA (2003) A mitochondrial DNA sequence variant associated with schizophrenia and oxidative stress. *Schizophr Res* 65: 33-8
- Margulis L (1976) Origins of Metabolic Systems. *Science* 192: 249
- Martorell L, Segues T, Folch G, Valero J, Joven J, Labad A, Vilella E (2006) New variants in the mitochondrial genomes of schizophrenic patients. *Eur J Hum Genet* 14: 520-8
- Mathews CK, Van Holde KE, Ahern KG (2002) *Bioquímica, Tercera edn.* Pearsons Addison Weasley, Madrid
- Maxam AM, Gilbert W (1977) A new method for sequencing DNA. *Proc Natl Acad Sci U S A* 74: 560-4
- McFarland R, Elson JL, Taylor RW, Howell N, Turnbull DM (2004) Assigning pathogenicity to mitochondrial tRNA mutations: when "definitely maybe" is not good enough. *Trends Genet* 20: 591-6
- Mishmar D, Ruiz-Pesini E, Golik P, Macaulay V, Clark AG, Hosseini S, Brandon M, Easley K, Chen E, Brown MD, Sukernik RI, Olckers A, Wallace DC (2003) Natural selection shaped regional mtDNA variation in humans. *Proc Natl Acad Sci U S A* 100: 171-6

- Mitchell AL, Elson JL, Howell N, Taylor RW, Turnbull DM (2006) Sequence variation in mitochondrial complex I genes: mutation or polymorphism? *J Med Genet* 43: 175-9
- Monnat RJ, Jr., Reay DT (1986) Nucleotide sequence identity of mitochondrial DNA from different human tissues. *Gene* 43: 205-11
- Mosquera-Miguel A, Álvarez-Iglesias V, Cerezo M, Lareu MV, Carracedo A, Salas A (2009) Testing the performance of mtSNP minisequencing in forensic samples. *Forensic Sci Int Genet* 3: 261-4
- Mullis K, Faloona F, Scharf S, Saiki R, Horn G, Erlich H (1986) Specific enzymatic amplification of DNA in vitro: the polymerase chain reaction. *Cold Spring Harb Symp Quant Biol* 51 Pt 1: 263-73
- Musgrave-Brown E, Ballard D, Balogh K, Bender K, Berger B, Bogus M, Borsting C, Brion M, Fondevila M, Harrison C, Oguzturun C, Parson W, Phillips C, Proff C, Ramos-Luis E, Sanchez JJ, Sanchez Diz P, Sobrino Rey B, Stradmann-Bellinghausen B, Thacker C, Carracedo A, Morling N, Scheithauer R, Schneider PM, Syndercombe Court D (2007) Forensic validation of the SNPforID 52-plex assay. *Forensic Sci Int Genet* 1: 186-90
- Oliveira G, Diogo L, Grazina M, Garcia P, Ataíde A, Marques C, Miguel T, Borges L, Vicente AM, Oliveira CR (2005) Mitochondrial dysfunction in autism spectrum disorders: a population-based study. *Dev Med Child Neurol* 47: 185-9
- Orita M, Iwahana H, Kanazawa H, Hayashi K, Sekiya T (1989) Detection of polymorphisms of human DNA by gel electrophoresis as single-strand conformation polymorphisms. *Proc Natl Acad Sci U S A* 86: 2766-70
- Paneto GG, Martins JA, Longo LV, Pereira GA, Freschi A, Alvarenga VL, Chen B, Oliveira RN, Hirata MH, Cicarelli RM (2007) Heteroplasmy in hair: differences among hair and blood from the same individuals are still a matter of debate. *Forensic Sci Int* 173: 117-21
- Parr RL, Maki J, Reguly B, Dakubo GD, Aguirre A, Wittock R, Robinson K, Jakupciak JP, Thayer RE (2006) The pseudo-mitochondrial genome influences mistakes in heteroplasmy interpretation. *BMC Genomics* 7: 185
- Parson W, Bandelt H-J (2007) Extended guidelines for mtDNA typing of population data in forensic science. *Forensic Sci Int Genet* 1: 13-9

- Parsons TJ, Muniec DS, Sullivan K, Woodyatt N, Alliston-Greiner R, Wilson MR, Berry DL, Holland KA, Weedn VW, Gill P, Holland MM (1997) A high observed substitution rate in the human mitochondrial DNA control region. *Nat Genet* 15: 363-8
- Perego UA, Achilli A, Angerhofer N, Accetturo M, Pala M, Olivieri A, Kashani BH, Ritchie KH, Scozzari R, Kong QP, Myres NM, Salas A, Semino O, Bandelt HJ, Woodward SR, Torroni A (2009) Distinctive Paleo-Indian migration routes from Beringia marked by two rare mtDNA haplogroups. *Curr Biol* 19: 1-8
- Pesole G, Gissi C, De Chirico A, Saccone C (1999) Nucleotide substitution rate of mammalian mitochondrial genomes. *J Mol Evol* 48: 427-34
- Pezzotti A, Kraft P, Hankinson SE, Hunter DJ, Buring J, Cox DG (2009) The mitochondrial A10398G polymorphism, interaction with alcohol consumption, and breast cancer risk. *PLoS One* 4: e5356
- Phillips C, Salas A, Sanchez JJ, Fondevila M, Gomez-Tato A, Álvarez-Dios J, Calaza M, de Cal MC, Ballard D, Lareu MV, Carracedo A (2007) Inferring ancestral origin using a single multiplex assay of ancestry-informative marker SNPs. *Forensic Sci Int Genet* 1: 273-80
- Pons R, Andreu AL, Checcarelli N, Vila MR, Engelstad K, Sue CM, Shungu D, Haggerty R, de Vivo DC, DiMauro S (2004) Mitochondrial DNA abnormalities and autistic spectrum disorders. *J Pediatr* 144: 81-5
- Pyle A, Foltynie T, Tiangyou W, Lambert C, Keers SM, Allcock LM, Davison J, Lewis SJ, Perry RH, Barker R, Burn DJ, Chinnery PF (2005) Mitochondrial DNA haplogroup cluster UKJT reduces the risk of PD. *Ann Neurol* 57: 564-7
- Quintáns B, Álvarez-Iglesias V, Salas A, Phillips C, Lareu MV, Carracedo A (2004) Typing of mitochondrial DNA coding region SNPs of forensic and anthropological interest using SNaPshot minisequencing. *Forensic Sci Int* 140: 251-7
- Ramos A, Santos C, Alvarez L, Nogues R, Aluja MP (2009) Human mitochondrial DNA complete amplification and sequencing: a new validated primer set that prevents nuclear DNA sequences of mitochondrial origin co-amplification. *Electrophoresis* 30: 1587-93

- Raule N, Sevini F, Santoro A, Altília S, Franceschi C (2007) Association studies on human mitochondrial DNA: methodological aspects and results in the most common age-related diseases. *Mitochondrion* 7: 29-38
- Richards MB, Macaulay VA, Bandelt H-J, Sykes BC (1998) Phylogeography of mitochondrial DNA in western Europe. *Ann Hum Genet* 62: 241-60
- Rieder MJ, Taylor SL, Tobe VO, Nickerson DA (1998) Automating the identification of DNA variations using quality-based fluorescence re-sequencing: analysis of the human mitochondrial genome. *Nucleic Acids Res* 26: 967-73
- Roostalu U, Kutuev I, Loogvali EL, Metspalu E, Tambets K, Reidla M, Khusnutdinova EK, Usanga E, Kivisild T, Villems R (2007) Origin and expansion of haplogroup H, the dominant human mitochondrial DNA lineage in West Eurasia: the Near Eastern and Caucasian perspective. *Mol Biol Evol* 24: 436-48
- Rosenberg NA, Pritchard JK, Weber JL, Cann HM, Kidd KK, Zhivotovsky LA, Feldman MW (2002) Genetic structure of human populations. *Science* 298: 2381-5
- Ross OA, McCormack R, Maxwell LD, Duguid RA, Quinn DJ, Barnett YA, Rea IM, El-Agnaf OM, Gibson JM, Wallace A, Middleton D, Curran MD (2003) mt4216C variant in linkage with the mtDNA TJ cluster may confer a susceptibility to mitochondrial dysfunction resulting in an increased risk of Parkinson's disease in the Irish. *Exp Gerontol* 38: 397-405
- Salas A, Bandelt H-J, Macaulay V, Richards MB (2007) Phylogeographic investigations: the role of trees in forensic genetics. *Forensic Sci Int* 168: 1-13
- Salas A, Carracedo A, Macaulay V, Richards M, Bandelt HJ (2005a) A practical guide to mitochondrial DNA error prevention in clinical, forensic, and population genetics. *Biochem Biophys Res Commun* 335: 891-9
- Salas A, Comas D, Lareu MV, Bertranpetit J, Carracedo A (1998) mtDNA analysis of the Galician population: a genetic edge of European variation. *Eur J Hum Genet* 6: 365-75
- Salas A, Fachal L, Marcos-Alonso S, Vega A, Martinon-Torres F (2009) Investigating the role of mitochondrial haplogroups in genetic predisposition to meningococcal disease. *PLoS One* 4: e8347

- Salas A, Lareu MV, Carracedo A (2001a) Heteroplasmy in mtDNA and the weight of evidence in forensic mtDNA analysis: a case report. *Int J Legal Med* 114: 186-90
- Salas A, Rasmussen EM, Lareu MV, Morling N, Carracedo A (2001b) Fluorescent SSCP of overlapping fragments (FSSCP-OF): a highly sensitive method for the screening of mitochondrial DNA variation. *Forensic Sci Int* 124: 97-103
- Salas A, Yao Y-G, Macaulay V, Vega A, Carracedo A, Bandelt H-J (2005b) A critical reassessment of the role of mitochondria in tumorigenesis. *PLoS Med* 2: e296
- Samuels DC, Carothers AD, Horton R, Chinnery PF (2006) The power to detect disease associations with mitochondrial DNA haplogroups. *Am J Hum Genet* 78: 713-20
- Sánchez JJ, Phillips C, Borsting C, Balogh K, Bogus M, Fondevila M, Harrison CD, Musgrave-Brown E, Salas A, Syndercombe-Court D, Schneider PM, Carracedo A, Morling N (2006) A multiplex assay with 52 single nucleotide polymorphisms for human identification. *Electrophoresis* 27: 1713-24
- Sanger F, Nicklen S, Coulson AR (1977) DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci U S A* 74: 5463-7
- Santos C, Montiel R, Sierra B, Bettencourt C, Fernandez E, Alvarez L, Lima M, Abade A, Aluja MP (2005) Understanding differences between phylogenetic and pedigree-derived mtDNA mutation rate: a model using families from the Azores Islands (Portugal). *Mol Biol Evol* 22: 1490-505
- Schwartz M, Vissing J (2002) Paternal inheritance of mitochondrial DNA. *N Engl J Med* 347: 576-80
- Schwartz M, Vissing J (2003) New patterns of inheritance in mitochondrial disease. *Biochem Biophys Res Commun* 310: 247-51
- Schwartz M, Vissing J (2004) No evidence for paternal inheritance of mtDNA in patients with sporadic mtDNA mutations. *J Neurol Sci* 218: 99-101
- Setiawan VW, Chu LH, John EM, Ding YC, Ingles SA, Bernstein L, Press MF, Ursin G, Haiman CA, Neuhausen SL (2008) Mitochondrial DNA G10398A variant is not associated with breast cancer in African-American women. *Cancer Genet Cytogenet* 181: 16-9

- Sigurgardóttir S, Helgason A, Gulcher JR, Stefansson K, Donnelly P (2000) The mutation rate in the human mtDNA control region. *Am J Hum Genet* 66: 1599-609
- Soares P, Ermini L, Thomson N, Mormina M, Rito T, Rohl A, Salas A, Oppenheimer S, Macaulay V, Richards MB (2009) Correcting for purifying selection: an improved human mitochondrial molecular clock. *Am J Hum Genet* 84: 740-59
- Sobrino B, Brión M, Carracedo A (2005) SNPs in forensic genetics: a review on SNP typing methodologies. *Forensic Sci Int* 154: 181-94
- Sullivan KM, Hopgood R, Lang B, Gill P (1991) Automated amplification and sequencing of human mitochondrial DNA. *Electrophoresis* 12: 17-21
- Sutovsky P, Moreno RD, Ramalho-Santos J, Dominko T, Simerly C, Schatten G (2000) Ubiquitinated sperm mitochondria, selective proteolysis, and the regulation of mitochondrial inheritance in mammalian embryos. *Biol Reprod* 63: 582-90
- Szibor R, Plate I, Heinrich M, Michael M, Schoning R, Wittig H, Lutz-Bonengel S (2007) Mitochondrial D-loop (CA)<sub>n</sub> repeat length heteroplasmy: frequency in a German population sample and inheritance studies in two pedigrees. *Int J Legal Med* 121: 207-13
- Taylor RW, McDonnell MT, Blakely EL, Chinnery PF, Taylor GA, Howell N, Zeviani M, Briem E, Carrara F, Turnbull DM (2003) Genotypes from patients indicate no paternal mitochondrial DNA contribution. *Ann Neurol* 54: 521-4
- Templeton AR (1997) Out of Africa? What do genes tell us? *Curr Opin Genet Dev* 7: 841-7
- Torrioni A, Achilli A, Macaulay V, Richards M, Bandelt HJ (2006) Harvesting the fruit of the human mtDNA tree. *Trends Genet* 22: 339-45
- Torrioni A, Huoponen K, Francalacci P, Petrozzi M, Morelli L, Scozzari R, Obinu D, Savontaus ML, Wallace DC (1996) Classification of European mtDNAs from an analysis of three European populations. *Genetics* 144: 1835-50
- Torrioni A, Neel JV, Barrantes R, Schurr TG, Wallace DC (1994) Mitochondrial DNA "clock" for the Amerinds and its implications for timing their entry into North America. *Proc Natl Acad Sci U S A* 91: 1158-62

- Torrioni A, Rengo C, Guida V, Cruciani F, Sellitto D, Coppa A, Calderon FL, Simionati B, Valle G, Richards M, Macaulay V, Scozzari R (2001) Do the four clades of the mtDNA haplogroup L2 evolve at different rates? *Am J Hum Genet* 69: 1348-56
- Torrioni A, Schurr TG, Yang CC, Szathmary EJ, Williams RC, Schanfield MS, Troup GA, Knowler WC, Lawrence DN, Weiss KM, et al. (1992) Native American mitochondrial DNA analysis indicates that the Amerind and the Nadene populations were founded by two independent migrations. *Genetics* 130: 153-62
- Tully G, Bär W, Brinkmann B, Carracedo A, Gill P, Morling N, Parson W, Schneider P (2001) Considerations by the European DNA profiling (EDNAP) group on the working practices, nomenclature and interpretation of mitochondrial DNA profiles. *Forensic Sci Int* 124: 83-91
- Tully G, Barritt SM, Bender K, Brignon E, Capelli C, Dima-Simonin N, Eichmann C, Ernst CM, Lambert C, Lareu MV, Ludes B, Mevag B, Parson W, Pfeiffer H, Salas A, Schneider PM, Staalstrom E (2004) Results of a collaborative study of the EDNAP group regarding mitochondrial DNA heteroplasmy and segregation in hair shafts. *Forensic Sci Int* 140: 1-11
- Tully G, Sullivan KM, Nixon P, Stones RE, Gill P (1996) Rapid detection of mitochondrial sequence polymorphisms using multiplex solid-phase fluorescent minisequencing. *Genomics* 34: 107-13
- Tully LA, Parsons TJ, Steighner RJ, Holland MM, Marino MA, Prenger VL (2000) A sensitive denaturing gradient-Gel electrophoresis assay reveals a high frequency of heteroplasmy in hypervariable region 1 of the human mtDNA control region. *Am J Hum Genet* 67: 432-43
- Ueno H, Nishigaki Y, Kong Q-P, Fuku N, Kojima S, Iwata N, Ozaki N, Tanaka M (2009) Analysis of mitochondrial DNA variants in Japanese patients with schizophrenia. *Mitochondrion* 9: 385-93
- Underhill PA, Kivisild T (2007) Use of Y chromosome and mitochondrial DNA population structure in tracing human migrations. *Annu Rev Genet* 41: 539-64
- Vallone PM, Just RS, Coble MD, Butler JM, Parsons TJ (2004) A multiplex allele-specific primer extension assay for forensically informative SNPs



- distributed throughout the mitochondrial genome. *Int J Legal Med* 118: 147-57
- van der Walt JM, Dementieva YA, Martin ER, Scott WK, Nicodemus KK, Kroner CC, Welsh-Bohmer KA, Saunders AM, Roses AD, Small GW, Schmechel DE, Murali Doraiswamy P, Gilbert JR, Haines JL, Vance JM, Pericak-Vance MA (2004) Analysis of European mitochondrial haplogroups with Alzheimer disease risk. *Neurosci Lett* 365: 28-32
- Vega A, Salas A, Milne RL, Carracedo B, Ribas G, Ruibal A, de León AC, Gonzalez-Hernandez A, Benitez J, Carracedo A (2009) Evaluating new candidate SNPs as low penetrance risk factors in sporadic breast cancer: a two-stage Spanish case-control study. *Gynecol Oncol* 112: 210-4
- Vigilant L, Pennington R, Harpending H, Kocher TD, Wilson AC (1989) Mitochondrial DNA sequences in single hairs from a southern African population. *Proc Natl Acad Sci U S A* 86: 9350-4
- Vigilant L, Stoneking M, Harpending H, Hawkes K, Wilson AC (1991) African populations and the evolution of human mitochondrial DNA. *Science* 253: 1503-7
- Wang X (2001) The expanding role of mitochondria in apoptosis. *Genes Dev* 15: 2922-33
- Ward RH, Frazier BL, Dew-Jager K, Paabo S (1991) Extensive mitochondrial diversity within a single Amerindian tribe. *Proc Natl Acad Sci U S A* 88: 8720-4
- Weissman JR, Kelley RI, Bauman ML, Cohen BH, Murray KF, Mitchell RL, Kern RL, Natowicz MR (2008) Mitochondrial disease in autism spectrum disorder patients: a cohort analysis. *PLoS One* 3: e3815
- White DJ, Gemmell NJ (2009) Can indirect tests detect a known recombination event in human mtDNA? *Mol Biol Evol* 26: 1435-9
- White HE, Durston VJ, Seller A, Fratter C, Harvey JF, Cross NC (2005) Accurate detection and quantitation of heteroplasmic mitochondrial point mutations by pyrosequencing. *Genet Test* 9: 190-9
- Wilson MR, Allard MW, Monson K, Miller KW, Budowle B (2002) Recommendations for consistent treatment of length variants in the human mitochondrial DNA control region. *Forensic Sci Int* 129: 35-42

- Wolpoff MH, Hawks J, Caspari R (2000) Multiregional, not multiple origins. *Am J Phys Anthropol* 112: 129-36
- Wyman AR, White R (1980) A highly polymorphic locus in human DNA. *Proc Natl Acad Sci U S A* 77: 6754-8
- Yao Y-G, Kong Q-P, Salas A, Bandelt H-J (2008) Pseudomitochondrial genome haunts disease studies. *J Med Genet* 45: 769-72
- Yao Y-G, Salas A, Bravi CM, Bandelt H-J (2006) A reappraisal of complete mtDNA variation in East Asian families with hearing impairment. *Hum Genet* 119: 505-15

## **Bibliografía adicional**

**Álvarez-Iglesias V**, Estudio multidisciplinar de la variabilidad del ADN mitocondrial en poblaciones humanas. Tesis doctoral, Universidade de Santiago de Compostela.